A STUDY IN THE ANALYSIS OF
STATIONARY TIME SERIES

BY

HERMAN WOLD

SECOND EDITION

WITH AN APPENDIX BY

PETER WHITTLE

ALMQVIST & WIKSELL

STOCKHOLM

FIRST EDITION 1938

PRINTED IN SWEDEN BY

Almqvist & Wiksells

BOKTRYCKERI AKTIEBOLAG

UPPSALA 1954

PREFACE.

In a sequence of fundamental memoirs, G. UDNY YULE, the eminent English statistician, has proposed certain methods of time series analysis which are of an essentially wider scope than the classical methods used in the search for periodicities. The basis of the new methods is a concept of flexible periodicity which in an ideal case reduces to the classical, functionally rigid periodicity. The importance and the broad applicability of the new ideas has been stressed particularly in subsequent discussion of the nature of business cycles.

In the recent rapid development of the theory of probability, the production of A. KHINTCHINE and A. KOLMOGOROFF represents a genuine discontinuity. A firm, axiomatic foundation has been obtained for the theory; other important results belong to the theory of random processes, i. e. hypothetical models for the analysis of time series. In accordance with the great diversity of time series, the main types of random process are of quite different structure.

In the theory of probability, the approaches of G. U. YULE fall under the heading of the stationary random process as defined and studied by A. KHINTCHINE. The present work might be described as a trial to subject the fertile methods of empirical analysis proposed by YULE to an examination and a development by the use of the mathematically strict tools supplied by the modern theory of probability. This statement, however, implies no valuation of the results and should be regarded rather as a tribute to my sources of inspiration and to the traditions of my milieu of study.

My most sincere thanks are due to my teacher, Professor HARALD CRAMÉR. His brilliant courses, distinguished by a spirit of realism combined with penetrating logic, have laid the foundation for my further work. As far as the present thesis is concerned, this is true not only in general but also in respect to particular parts thereof, as indicated by the references to his 1933 course on Time Series Analysis. I wish to evidence my deep gratitude to Professor CRAMÉR also for the encouragement and interest shown me at all

times, and culminating in his detailed reading of the first version of the manuscript. Our subsequent discussions have caused a revision particularly of the treatment of questions of convergence in probability.

To the Royal Swedish Academy of Sciences I want to express my respectful gratitude for a generous grant covering a substantial part of the expenses for printing and numerical calculation.

I am greatly indebted to my friends and colleagues Mr. G. ELFVING, Mr. W. FELLER and Mr. O. LUNDBERG for numerous stimulating discussions and for having read the manuscript and corrected many errors. I have also profited to a great extent by consultations with a large number of research workers in the different fields touched upon in the thesis. These obligations are, however, too comprehensive and indefinite to be expressed in detail.

— —

— —

Stockholm, July 1938. *H. W.*

PREFACE TO THE SECOND EDITION.

Stationary processes having in the last 15 years been the subject of intensive research, important results have been obtained both regarding their theory and their many fruitful applications. In presenting a new edition of my thesis, the recent development is briefly dealt with in Appendices 1–2 (these replace Appendices A–B of the first edition, which were devoted to special topics), whereas the main text is left unaltered, except for a slight revision that makes use only of material available in 1938. I am greatly indebted to Dr. WHITTLE for writing Appendix 2, in which two main lines of progress are surveyed, viz. spectral theory and methods of statistical inference The short Appendix 1 comments, by way of numbered foot-notes, a few specific points in the main text.

The first edition had the dual form of an expository survey and a research report. It is hoped that the second edition may still serve as an introduction to the theory and the applications of stationary processes.

Uppsala, March 1953. *H. W.*

Contents.

List of Theorems.

List of Graphs.

Introduction.

1. Remarks on the scope of the study.

Observational series which describe phenomena changing with time may be roughly classified in two broad categories, viz. *evolutive* and *stationary*. In the former case, different sections of the time series are dissimilar in one or more respects. For instance, the sectional averages may be distinctly different, or some other structural property of the series may present variation. In the analysis of evolutive time series, absolute time plays a fundamental role, e.g. as the independent variable in a trend function, or as a fixed scale in studying the development of a phenomenon from an initial state of rest.

Stationary time series are unchanging in respect to their general structure. The fluctuations up and down in such a series may seem random or show tendencies to regularity — in any case, the character of the series is, on the whole, the same in different sections. Or otherwise expressed, in the analysis of stationary series time is allotted the secondary role of a passive medium. Even without preparation, observational time series are frequently stationary. On the other hand, the deviations from a trend form a type of derived time series which is often stationary.

Among stationary as well as evolutive series we may distinguish a great many different types, and in actual fact the dissimilarities are deepgoing. As a consequence, the problems which are relevant from the viewpoint of the applications are different for the various types of series. Now regarding stationary series, if we judge from the earlier literature on the subject, their analysis might seem equivalent to the search for periodicities. In the present volume, let this be said at once, time series analysis is taken in a much wider sense.

Considering the classical methods of Fourier and Schuster, the hypothesis underlying these methods is that the time series under analysis might contain hidden periodicities, that is functional components which are periodic in the strict mathematical sense. It is well-known that these methods have often been applied with

definite success, and equally well-known that in many fields of scientific research they have met a severe criticism. An essential point in the criticism is that the idea of strict periods cannot possibly be realistic and adequate in certain applications. It has been claimed, for example, that in the theory of business cycles the hypothetical approach must be flexible to some extent, admit small changes in the periods and the amplitudes etc. A modified type of approach thus being called for, the difficulty is to find a precisely defined combination of the ideas of periodicity and of flexibility. It is evident that a strict hypothetical set-up as required can be reached only on the basis of the theory of probability.

Though the above mentioned critical argument is old, it was not until rather recently that approaches have been suggested which allow for changes in the waves in the time series under analysis. There are two main lines of approach, both of them germinating from G. U. YULE. Let these be briefly outlined.

Starting from a purely random series as given, for example, by dice-throwing, G. U. YULE ((1921), (1926)) forms the differences of a fixed order, and finds that the series thus obtained presents a tendency to regular fluctuations. E. SLUTSKY ((1927), (1937)) studies the effect of more general linear operations, and finds that under certain circumstances the resulting series will present sinusoidal waves with slowly changing amplitude and phase, waves showing a puzzling likeness to the cycles in economic time series. Nice examples of this are given by suitable moving averages of the primary random series. In the terminology of this study, the approaches mentioned are special cases of the scheme of *moving averages.*

The second type of approach is introduced by G. U. YULE (1927) in a study on sunspot numbers. Considering the sunspot index in a set of equidistant time points, YULE investigates the multiple correlation between these observations, and approximates by the use of linear regression analysis each observation by a linear function of the next preceding ones. The scheme thus implicitly defined will be called the scheme of *linear autoregression.* Using a physical interpretation, YULE gives a suggestive illustration of the new idea — a pendulum subjected to a stream of random shocks will be ruled by a scheme of linear autoregression. To a certain extent, the movement of the pendulum will bear resemblance to a free swinging, but the random impulses will cause a continuous shift in amplitude and phase.

Using a comprehensive term, the two schemes mentioned will be called schemes of *linear regression.* It is seen to be a common feature of these schemes that a random element plays a fundamental, active role. This constitutes a distinct contrast to the scheme of *hidden periodicities* — as we shall call the hypothesis of strict periods — and makes the schemes of linear regression *a priori* plausible in several instances where the scheme of hidden periodicities has been criticized.[1]

From the viewpoint of the theory of probability, the schemes of linear regression are special cases of the *stationary random process* as defined and studied by A. KHINTCHINE ((1932)—(1934)). Let us discuss the situation in some detail.

Considering a phenomenon as described by an observational time series, let us fix arbitrarily a finite set of time points, say $(t) = (t_1, t_2, \ldots, t_n)$. In a probabilistic theory of the phenomenon, we must necessarily assume that the behaviour of the time series in the n points considered is ruled by a definite probability distribution in n dimensions. Generally speaking, this distribution may be taken to be defined by a distribution function, say $F(t_1, \ldots, t_n; u_1, \ldots, u_n)$, where $u_1, \ldots, u_n$ are real variables. For instance, considering a set consisting of but one time point (t_1), the hypothetical function $F(t_1; u_1)$ will indicate the probability that the observational value in t_1 is less than or equal to u_1. Having stated this, it is clear that we must assume certain relations of consistency between the functions F which belong to different sets (t); otherwise the hypothetical set-up might contradict itself. For instance, it is evident that $F(t_1; u_1)$ must be assumed to satisfy all relations of the type $F(t_1; u_1) = F(t_1, t_2; u_1, \infty)$.

We have seen that the probabilistic treatment of a time series requires a set of distribution functions, say $\{F\}$, such that there is one function F corresponding to every finite set (t) of time points, and that the functions F satisfy certain consistency relations. On the other hand, such a hypothetical set-up will give a sufficient basis for a formal probabilistic analysis. Any set $\{F\}$ with properties as mentioned is called a *random process*, and according to a fundamental theorem of A. KOLMOGOROFF (1933) such a set $\{F\}$ is equivalent to a probability distribution in an infinite number of dimensions. Of course, each time point corresponds to one dimension in this distribution.

In defining a random process, we may either choose our points

(t) quite freely or else restrict them somehow, e. g. to an unbroken sequence of equidistant time points. Again following A. KOLMOGOROFF, a random process as defined by the set $\{F\}$ is in the first case called *continuous*, in the case of equidistance *discrete*. Interpreting the random processes as random variables in an infinite number of dimensions, it is seen that the number of dimensions is enumerative in the case of a discrete process, and non-enumerative in the continuous case.

In applying a probabilistic scheme to a statistical distribution — perhaps multidimensional — each individual observation is looked upon as a sample value belonging to the hypothetical distribution in one or more dimensions. In the same manner in interpreting an observational time series as belonging to a random process the series is regarded as a sample value of the corresponding distribution in an infinite number of dimensions. Since a whole time series thus constitutes but one element in the statistical population, it is evident that the unspecified random process is a concept which is far too general to be useful in practical applications. According to the structure of the time series under analysis, we have to apply special types of the random process.

A. KHINTCHINE's definition of the stationary random process runs as follows. Letting $(t) = (t_1, t_2, \ldots, t_n)$ represent an arbitrary set of time points, and fixing arbitrarily a translation in time of this set, say $(t^*) = (t_1 + t,\ t_2 + t, \ldots, t_n + t)$, a random process as defined by a set $\{F\}$ of distribution functions is called stationary if the two functions F belonging to the two sets t and t^* are identical. Thus, the probability laws assumed to rule the observational time series depend on time in such a way that if we replace time as measured from a fixed time point by a time variable measured from another time point, the probability laws will remain the same. In other words, if the development in a time series is known up to a certain time point, say t, the probability laws ruling the continued development will depend only on the behaviour of the time series up to the time point t, not on the actual value of t. It can be seen that the postulated relativity of time is precisely adequate in view of the broad class of stationary time series roughly characterized above.

The present study is exclusively concerned with the theory and the applications of the *discrete stationary process*. Accordingly, problems concerning evolutive time series fall outside the program. For instance, trend analysis will not be dealt with at all. Further,

the case of continuous stationary processes will be touched upon only incidentally.

As far as I know, A. KHINTCHINE is alone in having dealt with the discrete stationary process in full generality; his chief result is that the stationary processes are ruled by a law of great numbers. The present study being concerned with other aspects of the stationary process, we shall next give a few comments on the main lines followed.

Chapter I serves a double purpose. In surveying the leading methods in the search for periodicities, particular attention is drawn to the hypotheses underlying these methods. By thus pointing out the rather narrow basis of the methods considered, the need for other types of hypothetical scheme is made clear, and the analysis of other types of approach prepared. On the other hand, after this rather detailed survey, the hypothesis of hidden periodicities will need no further separate treatment.

Chapter II is reserved for a general analysis of the discrete stationary process. Sections 13 and 14 are preparatory, and show that in certain respects the stationary processes may be dealt with in the same manner as random variables in a finite number of dimensions. In particular, the singular case introduced in section 14 corresponds to a multi-dimensional probability distribution which is entirely concentrated in a plane or some other linear sub-space.

The discrete stationary process being extremely general, sections 15 and 16 give a few examples of processes obtained by different specializations. In this way we arrive at strict definitions of the processes of linear regression which cover the above mentioned approaches suggested by G. U. YULE. Further, the scheme of hidden periodicities is obtained by means of a singular stationary process. Detailed illustrations of the processes thus defined are given through model time series, i. e. series constructed in an artificial way on the basis of random sampling numbers. Finally, in section 16, the normal stationary process is defined by a straightforward generalization of the normal distribution in a finite number of dimensions.

In the latter part of the second chapter, the structural properties of the general stationary process are studied. The field being wide and unexplored, the analysis has been restricted to the elementary features. Generally speaking, only such properties have been taken into consideration which could be studied by the use of linear

operations and autocorrelation coefficients. As the autocorrelation coefficients correspond to the mixed moments of second order in a set of ordinary, one-dimensional variables, the analysis will, in certain respects, be parallel to the familiar theory of multiple correlation.

In section 17 is shown that the autocorrelation coefficients of a discrete stationary process may always be interpreted as the FOURIER coefficients of a non-decreasing function. This theorem discloses clearly that it is only in special cases that a periodogram can tell us something relevant about a time series.

The linear autoregression analysis as prepared in section 18 and developed in section 19 is based on the idea of subjecting the discrete stationary process to a treatment which is parallel to a time series analysis by means of the methods proposed by G. U. YULE (1927) and already referred to. The autoregression analysis thus corresponds to a linear regression analysis in a finite set of one-dimensional variables. The periodogram analysis, on the other hand, may be interpreted as a graduation by means of simple harmonics. If we consider the forecasts delivered by the two methods, the autoregression analysis reaches the limit beyond which we cannot proceed when employing only linear methods.

Using the same tools as in section 19, the analysis of the structure of the discrete stationary process is, in section 20, carried further in a quite new direction. In spite of its wide comprehensiveness, the general discrete stationary process is found to be of a readily surveyable structure. In fact, the general process is built up by two components which may be interpreted as generalized processes of hidden periodicities and of linear regression respectively.

In point of principle, the methods used in Chapter II are of a scope which would admit of generalizations of the analysis in different directions. Using non-linear operations, it would be possible to perform an autoregression analysis corresponding to curvilinear regression analysis in the case of a finite number of variables. Further, the analysis might be extended to the case when the time series is multi-dimensional, i. e. when several properties of the phenomenon observed are studied simultaneously.

The stochastical difference equations studied in the first two sections of Chapter III form a generalization of the ordinary linear difference equations. While the solutions of the latter equations are ordinary functions, the solutions of the former equations are

discrete random processes. Considering an ordinary linear difference equation, its solution describes how a certain oscillatory mechanism will develop under given conditions. On the other hand, if a primary stream of impulses is defined by means of a random process, the solutions of a corresponding stochastical difference equation give the probability laws which will rule the oscillatory mechanism when subjected to the primary impulses. Of course, if the actual development of the mechanism is known, the corresponding series of impulses is readily obtained. Moreover, among the solutions of stochastical difference equations we find both evolutive and stationary random processes, e. g. the process of linear autoregression as strictly defined in section 15.

The latter part of Chapter III is reserved for a detailed study of the processes of linear regression, a chief purpose being to illustrate the general analysis in Chapter II, and the theory of stochastical difference equations. Particular attention is paid to the forecast situation. In contradistinction to the processes of hidden periodicities, the processes of linear regression contain an active random element which affects the efficiency of the forecast. As a matter of fact, in a process of linear regression, the efficiency is the less, the longer the interval of time forecasted over. On the other hand, the short time forecasts are the more efficient. In view of the applications this circumstance is advantageous, for, of course, the main interest is always focussed upon the short time forecasts.

The analysis in Chapter III, too, suggests certain generalizations. Thus, nothing prevents us from defining random processes by means of non-linear stochastical difference equations. Moreover, multi-dimensional random processes may be defined on the basis of systems of stochastical difference relations. A few remarks along the latter line are given in sections 31 and 32.

Chapter IV, finally, gives a few applications of the theoretical analysis to observational time series of the stationary type. Such a series being given, its correlogram — my term for the auto-correlation periodogram — is adopted as an indicator of which type of process to apply. Explicit applications being given of the processes of linear autoregression and of moving averages, general methods are indicated for finding suitable numerical values of the parameters involved. For instance, assuming a given time series to be a moving average of an unknown primary series, suitable values for the weights of the hypothetical moving average are derived

from the correlogram of the given series. Further, a general method is given for deriving the primary series which corresponds to such a set of weights.

The purpose of the applications is to illustrate certain general methods of analysis, not to supply a theory of the phenomena described by the time series examined. Accordingly, I make little account of the hypothetical schemes arrived at, and no attempts to test the significance of the parameters determined. On the contrary, warnings are repeatedly given for attaching importance to the numerical results of the analysis, for one reason because significance questions are extremely intricate in time series analysis. (It has turned out, however, that the results find support by the test methods that have been established after the 1st edition of this book was published. In Appendix 2 of this 2nd edition, P. Whittle has kindly undertaken to summarize the general test methods he has developed.)

2. **Principles of notation.**

The present volume being concerned with both theory and applications, symbols are needed for probabilistic as well as statistical concepts. Since the purpose of this section is to indicate the principles of notation, no completeness is aimed at with respect to the definitions of the elementary concepts considered. For these, reference is made to G. U. Yule and M. G. Kendall (1937) as to the theory of statistics, and to H. Cramér (1937) as to the theory of probabilities.

Generally speaking, the notations try to bring into relief the correspondence between theoretical and empirical concepts. Thus, for parallel theoretical and empirical concepts the same symbol will be used — in the latter case marked by a bar. As a rule, Greek letters will be reserved for random variables, Roman letters for functions, ordinary variables, and constants.

In agreement with these general principles, random variables as dealt with in statistics will be denoted by $\bar{\xi}$, $\bar{\eta}$, etc. Considering such a random or statistical variable, say $\bar{\xi}$, its observational values in a particular statistical population will be distinguished by running indices, e. g.

$$(1) \qquad \bar{\xi}_1, \bar{\xi}_2, \ldots, \bar{\xi}_n.$$

Irrespective of the order of the elements, an observational set (1) is uniquely determined by the corresponding function of cumulative relative frequencies, say $\bar{F}(u)$. Thus $\bar{F}(u)$, for every real u, equals the relative number of statistical units with an observational value $\bar{\xi}_i \leq u$. The function $\bar{F}(u)$ will be called the (empirical) distribution function of the observational set. By the use of the STIELTJES integral, the elementary characteristics of an empirical distribution (1) can be conveniently expressed in terms of $\bar{F}(u)$. The average of $\bar{\xi}$ in the population considered is

$$\bar{m} = \frac{1}{n} \cdot \sum_{i=1}^{n} \bar{\xi}_i = \int_{-\infty}^{\infty} u \cdot d\,\bar{F}(u).$$

The central moment of order k is denoted by $\bar{\mu}_k$, and reads

$$\bar{\mu}_k = \frac{1}{n} \cdot \sum_{i=1}^{n} (\bar{\xi}_i - \bar{m})^k = \int_{-\infty}^{\infty} (u - \bar{m})^k \cdot d\,\bar{F}(u).$$

Thus, $\bar{\mu}_2$ is the variance. Denoting the dispersion by $\bar{D}$, we have

(2) $$\bar{D}^2 = \bar{\mu}_2.$$

In studying a statistical variable $\bar{\xi}$, we introduce a corresponding hypothetical random variable ξ. Such a variable ξ, which is also called stochastical or aleatory, is completely characterized by its distribution function $F(u)$. By definition, $F(u)$ indicates the probability that ξ is less than or equal to u. This is expressed by the relation

(3) $$F(u) = P\{\xi \leq u\},$$

where P is the probability function of ξ.

In the analysis of a variable ξ, it is actually only the observational values $\bar{\xi}_i$, and the variable ξ as defined by the distribution function $F(u)$, which will appear in the formal developments. However, in the literature the observational values $\bar{\xi}_i$ are also often supplied with hypothetical parallels, called sample values. For instance, the distribution of ξ is spoken of as constituted by an infinite population of sample values ξ_i. Such a terminology is often convenient, and is, incidentally, used also in the present study. As we need no notation for these sample values, the symbol ξ_i will,

as indicated later on, be reserved for representing independent variables in random sampling.

Let $g(x)$ be a function of a real variable, and ξ a random variable with distribution function $F(u)$. Under general conditions, $g(\xi)$ represents a random variable with expectation given by

$$E[g(\xi)] = \int_{-\infty}^{\infty} g(u) \cdot dF(u). \tag{4}$$

In particular, the elementary characteristics of ξ may be interpreted in this way. For instance, the mean (m), dispersion (D), and the central moments (μ) are given by

$$m = E[\xi] = \int_{-\infty}^{\infty} u \cdot dF(u), \qquad D^2(\xi) = \mu_2. \tag{5}$$

$$\mu_k = E[(\xi - m)^k] = \int_{-\infty}^{\infty} (u - m)^k \cdot dF(u). \tag{6}$$

Any multi-dimensional random variable may be looked upon as a combination of one-dimensional variables. For such variables we shall use notations of type $\xi = [\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(h)}]$. Interpreting in this manner, we can often use the same notations as in the one-dimensional case. When full information is required, e. g. concerning (3), we shall use notations such as

$$F(u_1, u_2, \ldots, u_h) = P[\xi^{(1)} \leq u_1, \xi^{(2)} \leq u_2, \ldots, \xi^{(h)} \leq u_h].$$

In the same way, the expression (4) may be regarded as the expectation of a function $g(\xi)$ of a multi-dimensional variable ξ, only that $E[g(\xi)]$ must be interpreted as a vector in the space of $g(\xi)$, and that the integral must be extended over the space of ξ. For instance, let a one-dimensional random variable $p(\xi)$ be defined as a function of an h-dimensional variable ξ with distribution function $F(u) = F(u_1, \ldots, u_h)$. Then the distribution function of $p(\xi)$, say $G_p(x)$, will equal, for an arbitrarily fixed x, the expectation of a function $g_{x,p}(u) = g_{x,p}(u_1, \ldots, u_h)$ defined by the relations $g_{x,p}(u) = 1$ as $p(u) \leq x$, and $g_{x,p}(u) = 0$ as $p(u) > x$. Thus,

$$G_p(x) = E[g_{x,p}(\xi)] = \int_{R_h} g_{x,p}(u_1, \ldots, u_h) \cdot dF(u_1, \ldots, u_h).$$

In particular, letting $F(u_1, \ldots, u_h)$ be the distribution function of

a variable $\xi = [\xi^{(1)}, \ldots, \xi^{(h)}]$, and $G_p(x)$ that of the sum $p(\xi) = p_1 \xi^{(1)} + p_2 \xi^{(2)} + \cdots + p_h \xi^{(h)}$, where the p_i's are real constants, we have

$$(7) \qquad G_p(x) = \int_{p(x)} d_{u_1, \ldots, u_h} F(u_1, \ldots, u_h),$$

the integration being extended over a half-space $p(x)$ defined by $p(x) = [p_1 u_1 + \cdots + p_h u_h \leq x]$. The present study being chiefly concerned with linear functions of random variables, this formula will frequently come into use.

Considering a combined variable $\xi = [\xi^{(1)}, \ldots, \xi^{(h)}]$ with distribution function $F(u_1, \ldots, u_h)$, and taking $p_k = 0$ for $k \neq i$, and $p_i = 1$, formula (7) gives the distribution function of the individual variable $\xi^{(i)}$. Denoting the resulting functions by $F_i(u)$, the variables $\xi^{(i)}$ are called *independent* if, for an arbitrary set $(u_1, \ldots, u_h)$, the following relation is satisfied,

$$F(u_1, u_2, \ldots, u_h) = F_1(u_1) \cdot F_2(u_2) \ldots F_h(u_h).$$

In the case of independent variables, the expectation satisfies a general relation involving arbitrary functions g_i, viz.

$$(8) \quad E[g_1(\xi^{(1)}) \cdot g_2(\xi^{(2)}) \ldots g_h(\xi^{(h)})] = E[g_1(\xi^{(1)})] \cdot E[g_2(\xi^{(2)})] \ldots E[g_h(\xi^{(h)})].$$

Putting $p_i = 1$ for all i, formula (7) gives the distribution function, say $G(u)$, of the sum of our h variables $\xi^{(i)}$. If these are independent, $G(u)$ can be expressed by the use of the familiar *composition (convolution, Faltung) symbol* $*$, viz.

$$(9) \qquad G(u) = F_1(u) * F_2(u) * \ldots * F_h(u).$$

In the case of two independent variables, the convolution is given by

$$G(u) = F_1(u) * F_2(u) = \int_{-\infty}^{\infty} F_1(u - x) \cdot d F_2(x).$$

A *random sample* containing h elements, and belonging to a random variable ξ with distribution function $F(u)$, may be defined as a sample value of an h-dimensional variable obtained by combining h independent variables ξ_i which have the same distribution function $F(u)$. The concept of a random sample is thus purely theoretical, and corresponds to that of a statistical population. As already men-

tioned, we need no particular notation for the elements in a random sample.

The ordinary correlation coefficient between two interdependent random variables, say $\xi(i)$ and $\xi(k)$, will be denoted $r = r(\xi(i); \xi(k))$.

Let $\ldots, \xi(-1), \xi(0), \xi(1), \xi(2), \ldots$ be a sequence of interdependent random variables such that the characteristics

$$m(\xi;\, t) = E[\xi(t)]; \qquad \nu_2^{(k)}(\xi;\, t) = E[\xi(t) \cdot \xi(t+k)] \tag{10}$$

will exist for all integral t and k. Let it further be assumed that the quantities defined by

$$\begin{cases} m(\xi) = \lim\limits_{\substack{n\to\infty \\ n'\to-\infty}} \dfrac{1}{n-n'+1} \cdot \sum\limits_{t=n'}^{n} m(\xi;\, t); \\ \nu_2^{(k)}(\xi) = \lim\limits_{\substack{n\to\infty \\ n'\to-\infty}} \dfrac{1}{n-n'+1} \cdot \sum\limits_{t=n'}^{n} \nu_2^{(k)}(\xi;\, t) \end{cases} \tag{11}$$

will exist. Under these circumstances, the coefficients r_k defined by

$$r_k = r_k(\xi) = \frac{\nu_2^{(k)} - m^2}{\nu_2^{(0)} - m^2} \tag{12}$$

will be called the *autocorrelation coefficients* of the sequence $\xi(t)$. It should be observed that this definition holds also in the special case when $\xi(t)$ reduces to an ordinary function of t.

Let (1) represent an empirical time series obtained by observations in the equidistant time points $t = 1, 2, \ldots, n$. Following G. U. YULE (1926), the coefficients $\bar{r}_k$ defined by

$$\bar{r}_k = \frac{\sum\limits_{i=1}^{n-k} \bar{\xi}_i \cdot \bar{\xi}_{i+k} - (n-k) \cdot \bar{m}_1 \cdot \bar{m}_2}{(n-k) \cdot \bar{D}_1 \cdot \bar{D}_2}, \tag{13}$$

where

$$\bar{m}_1 = \frac{1}{n-k} \cdot \sum_{1}^{n-k} \bar{\xi}_i; \qquad \bar{m}_2 = \frac{1}{n-k} \cdot \sum_{k+1}^{n} \bar{\xi}_i,$$

and

$$\bar{D}_1^{\,2} = \frac{1}{n-k} \cdot \sum_{1}^{n-k} (\bar{\xi}_i - \bar{m}_1)^2; \qquad \bar{D}_2^{\,2} = \frac{1}{n-k} \cdot \sum_{k+1}^{n} (\bar{\xi}_i - \bar{m}_2)^2,$$

will be called *serial coefficients*. These obviously form an empirical parallel to the autocorrelation coefficients.

CHAPTER I.

A survey of hypotheses and methods proposed for the analysis of time series.

3. Scope and disposition.

This chapter aims at giving a historical perspective to the investigations in the subsequent chapters. Within the bounds of the survey fall the fundamental facts concerning the principal theoretical schemes set up for the study of stationary time series in one dimension. The leading methods for fitting the considered schemes to observational data will also be examined. For the sake of concreteness some descriptive methods will be touched upon incidentally. The survey being concerned with a general outline only, reference is given to H. BURKHARDT (1904) and K. STUMPFF ((1927), (1937)) for further material.

The purely functional schemes will be dealt with first among the theoretical models for a given time series. At the opposite extreme is taken that purely probabilistic scheme in which the series is regarded as a random sample of an aleatory variable. The other schemes may be looked upon as intermediate cases. Of the mixed schemes, the approach of hidden periodicities is treated first. The series is here assumed to be additively composed of independent functional and random elements. The last section of the survey deals with the scheme of moving averages, studied in certain special cases by G. U. YULE (1921) and E. SLUTSKY (1927), and with the scheme proposed by G. U. YULE (1927) under the name of disturbed harmonics (cf. p. 2).

The time points considered will be equidistant. The unit of equidistance will be taken for the time unit, a simplification which evidently does not involve any loss of generality.

4. Functional schemes.

The functional schemes aim at a perfect functional representation of the empirical data.

As a first example of functional approach, we take the hypothesis of a periodic function. Assuming the period to equal h, and denoting the hypothetical function by $x(t)$, the following relation will be satisfied for any t

$$x(t) - x(t-h) = 0. \tag{14}$$

As t is given for integral values only, it will be sufficient to consider the case of an integral h. Then (14) forms an ordinary difference equation of order h. According to the theory of difference equations, the solutions of (14) may be written

$$\begin{aligned} x(t) &= m + \sum_k C_k \cos\left(\frac{2\pi}{h}\cdot kt + \varphi_k\right) = \\ &= m + \sum_k \left(A_k \cos\frac{2\pi}{h}kt + B_k \sin\frac{2\pi}{h}kt\right), \end{aligned} \tag{15}$$

where k runs from 1 to $(h-1)/2$ or $h/2$. The real parameters A_k, B_k, $C_k > 0$, and φ_k in $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ are connected by the relations

$$C_k^2 = A_k^2 + B_k^2, \qquad \varphi_k = \operatorname{arc\,tg}\,(-B_k/A_k). \tag{16}$$

The approach now described will be termed the *scheme of periodic functions*. According to (15), the approach function may be considered to be composed of superposed harmonics, each having an amplitude C_k, a phase φ_k, and an angular frequency λ_k given by

$$\lambda_k = \frac{2\pi}{h}\cdot k.$$

The expression (15) may be taken as the basis for various generalized hypotheses. We start with the approach

$$\begin{aligned} x(t) &= m + \sum_{k=1}^{s} C_k \cos(\lambda_k t + \varphi_k) = \\ &= m + \sum_{k=1}^{s} (A_k \cos\lambda_k t + B_k \sin\lambda_k t), \end{aligned} \tag{17}$$

which will be called the *scheme of superposed harmonics*. Alternately, the function (17) will be called a *composed harmonic*.

Denoting by p_k the periods of the individual harmonics in (15) or (17), we have

$$p_k = 2\,\pi / \lambda_k. \tag{18}$$

In the scheme (15) the periods of the individual harmonics representing $x(t)$ are seen to be true fractions of h. In the generalized scheme (17) this restriction will not be laid down on the individual periods as given by (18). On the other hand, since t takes on integral values only, it follows that for any integer n the substitution of $\lambda_k - n \cdot 2\,\pi$ for λ_k will have no effect in (17). Neither will $x(t)$ be affected by a simultaneous substitution of $-\lambda_k$ for λ_k, and $-\varphi_k$ for φ_k. Thus it would involve no loss of generality to assume that $0 < \lambda_k \leq \pi$. This means that, in point of principle, the analysis must be restricted to periods not less than 2 time units. A study of shorter periods requires a reduction of the time unit chosen as a basis for the analysis. However, unless explicitly stated otherwise, we shall merely assume that $0 < \lambda_k$.

A function $x(t)$ of type (17) belongs, as is well known, to the class of almost periodic functions in the sense of H. Bohr (see e. g. (1925)). The following property of an almost periodic function $x(t)$ is recorded for later use: An $\varepsilon > 0$ being arbitrarily given, there exists for every number $t_0 > 0$ an integer $T(\varepsilon,\ t_0) > t_0$ such that for every t (see H. Bohr (1925), p. 88)

$$|x(t + T) - x(t)| < \varepsilon. \tag{19}$$

A rough description of a scheme of superposed harmonics is yielded by its periodogram. This has the frequency $\lambda > 0$ for horizontal axis, and indicates by ordinates in λ_k the corresponding squared amplitudes C_k^2. It is seen that the periodogram — which is met in several variants, e. g. with the periods $p_k = 2\,\pi / \lambda_k$ as abscissae or with the amplitudes C_k as ordinates — does not pay any regard to the constant term or to the phases appearing in the expression (17). Another variant is the integrated periodogram. This is a function, say $S(\lambda)$, defined by

$$S(\lambda) = \sum_{\lambda_k \leq \lambda} C_k^2 \Big/ \sum_{k=1}^{s} C_k^2. \tag{20}$$

Thus $S(\lambda)$ is a step (or *saltus*) function which is proportionate to the

sum of squared amplitudes with frequency not greater than λ. An example of periodogram and corresponding integrated periodogram is given in the figure below.

In the study of light, the periodogram has an experimental parallel in the spectrum. The prism spreads the light waves according to their frequency λ_k, and the individual lines in the spectrum indicate the energy of respective wave components. This energy is proportionate to the squared amplitude. The energy represented in an interval of the spectrum is thus proportionate to the sum of the

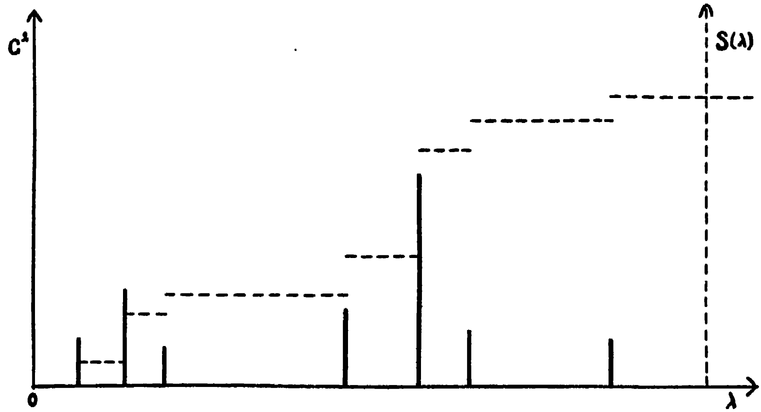

Fig. 1. Ordinary periodogram (C^2, vertical lines), and corresponding integrated periodogram ($S(\lambda)$, horizontal lines).

ordinates in the corresponding interval of the ordinary periodogram, and proportionate to the increase of the integrated periodogram in the same interval.

Analysis of white light produced spectra, where the lines were thin and lying very close together. This fact gave rise to the idea of continuous spectra and periodograms, in so far as the energy belonging to an interval was thought of as the integral of a spectral density. A survey and a development of the mathematical theory used in this connexion has been given by N. WIENER (1930). Translating to the terminology of the present study, this theory is based upon an analysis of the function

$$Q(u) = \lim_{z \to \infty} \frac{1}{2z} \cdot \int_{-z}^{z} \bigl(x(t+u) - m\bigr) \cdot \bigl(x(t) - m\bigr)\, dt, \tag{21}$$

where

$$m = \lim_{z \to \infty} \frac{1}{2z} \int_{-z}^{z} x(t)\, dt.$$

It is seen that if $x(t)$ is given by (17), $Q(u)$ reduces to

$$\frac{1}{2} \cdot \sum_{k=1}^{s} C_k^2 \cdot \cos \lambda_k u. \tag{22}$$

Modifying by a constant factor the function $S(\lambda)$ defined by WIENER, this reduces in the non-complex case to

(23)
$$S(\lambda)=\frac{2}{\pi}\cdot\int_0^\infty\frac{Q(u)}{Q(0)}\cdot\frac{\sin\lambda u}{u}\,du.$$

$S(\lambda)$ is called the *integrated periodogram* of $x(t)$, for in the case of superposed harmonics (17), $S(\lambda)$ as given by (23) reduces to (20) (see e. g. H. C. CARSLAW (1930), p. 322). WIENER shows that $S(\lambda)$ is always a non-decreasing function, and that there exist functions $x(t)$ with continuous integrated periodograms (23).

N. WIENER applies the generalized harmonic analysis also to functions defined by a random scheme. For instance, if $x(t)$ in integral intervals is 1 or -1 with equal probability, and if the values taken on in different intervals are independent, $S(\lambda)$ is with probability 1 given by

(24)
$$S(\lambda)=\frac{2}{\pi}\cdot\int_0^\lambda\frac{1-\cos u}{u^2}\,du.$$

5. On applied harmonic analysis.

Let an observational time series be represented by (1). If we wish to apply the scheme (17), the primary problem is to find for the parameters involved numerical values yielding as good fit as possible to the empirical data.

In case the observational data are strictly periodic, say with period p, an application of (15) by means of the FOURIER analysis will yield a perfect fit. It will be sufficient to consider the data $\bar{\xi}_1,\ldots,\bar{\xi}_p$ ranging over one period. However, since the formulae become particularly simple in case of an even period, and since we may take the double period for a basis if p is odd, let it be assumed that $h=p=2q$. The FOURIER formulae for the $2q$ parameters m, A_q, A_k, B_k, where $k=1, 2, \ldots, q-1$, then read as follows (see e.g. H. S. CARSLAW (1930), p. 325)

(25)
$$\begin{cases} m=\dfrac{1}{p}\cdot\sum\limits_{t=1}^{p}\bar{\xi}_t, & A_q=\dfrac{1}{p}\cdot\sum\limits_{t=1}^{p}\bar{\xi}_t\cos\pi t,\\[2ex] A_k=\dfrac{1}{q}\cdot\sum\limits_{t=1}^{p}\bar{\xi}_t\cos\dfrac{\pi}{q}kt, & B_k=\dfrac{1}{q}\cdot\sum\limits_{t=1}^{p}\bar{\xi}_t\sin\dfrac{\pi}{q}kt.\end{cases}$$

In the approach (17), the essential problem is to evaluate the frequency numbers λ_k. The principal method is that of A. SCHUSTER ((1898), (1900)) which is based upon the construction of an empirical

periodogram, say $\overline{C}^2(\lambda)$. The formulae required are (see e.g. K. Stumpff (1927), p. 103)

$$(26)\qquad \begin{cases} A(\lambda) = \dfrac{2}{n} \cdot \sum\limits_{t=1}^{n} (\bar{\xi}_t - \bar{m}) \cdot \cos \lambda t, & 0 < \lambda < \pi, \\ B(\lambda) = \dfrac{2}{n} \cdot \sum\limits_{t=1}^{n} (\bar{\xi}_t - \bar{m}) \cdot \sin \lambda t, & 0 < \lambda < \pi; \end{cases}$$

$$(27)\qquad \overline{C}^2(\lambda) = A^2(\lambda) + B^2(\lambda).$$

For $\lambda = \pi$, the factor 2 in A and B must be omitted.

A graph of the curve $\overline{C}^2(\lambda)$ presents characteristic maxima, the abscissae λ_k of which are taken for the frequency numbers sought for. The corresponding parameters $A_k = A(\lambda_k)$ and $B_k = B(\lambda_k)$ are obtained from (26).

The periodogram method gives much valuable information about the series under investigation, but is rather inconvenient. However, a careful fit of (17) to observational data seems to necessitate tedious computations, so the labour seems due to the problem, not to the method. Nevertheless, many simplified and, accordingly, approximate methods have been proposed. One type of these is of interest for the sequel because it is based upon the differential or difference relations satisfied by a sum of harmonics. The first method of this kind, that of S. Oppenheim (1909), utilizes the fact that for arbitrary C_k and φ_k the function $x(t)$ given by (17) satisfies a differential equation of $2s$:th order,

$$(28)\qquad x^{(2s)}(t) + g_1 \cdot x^{(2s-2)}(t) + \cdots + g_{s-1} \cdot x^{(2)}(t) + g_s \cdot [x(t) - m] = 0,$$

with constant coefficients g_i such that the equation

$$(29)\qquad z^s + g_1 \cdot z^{s-1} + \cdots + g_{s-1} \cdot z + g_s = 0$$

has the roots $-\lambda_1^2, -\lambda_2^2, \ldots, -\lambda_s^2$. Identifying $\bar{\xi}_t$ with $x(t)$, and taking $m = \bar{m}$, S. Oppenheim uses the identity (28) for s successive t-values, the g_i being so far undetermined. These relations are considered a system determining the g_i's. Inserting the resulting g_i's in (29), the solving of this equation gives, finally, the frequency numbers λ_k desired.

The derivatives required for the system of identities (28) S. Oppenheim obtains from the observational differences $\Delta^{2i}\,\bar{\xi}_t$ by the

well-known serial development (see e. g. E. T. WHITTAKER and G. ROBINSON (1926), p. 64). The intricate passage from differences to differentials is avoided in the modification of the OPPENHEIM method given by H. BRUNS (1911), who starts from a certain identity between central differences satisfied by $x(t)$, *viz.*

$$\Delta^{2s} x(t-s) + h_1 \cdot \Delta^{2s-2} x(t-s+1) + \cdots + h_{s-1} \cdot \Delta^2 x(t-1) + h_s [x(t) - m] = 0. \tag{30}$$

Here the constant coefficients h_i are such that the roots of the equation

$$z^s + h_1 \cdot z^{s-1} + \cdots + h_{s-1} \cdot z + h_s = 0$$

are $-q_k^2$, where

$$q_k = 2 \sin \lambda_k/2. \tag{31}$$

The fact that different functional schemes may give a good fit gives rise to the question of which scheme should be preferred when analysing a given time series. This is a particular aspect of the general test problem which is fundamental in all practical applications. However, the most important aspects of the test problem belong to the theory of probability, and will be touched upon later. As purely functional test methods may be regarded those which in principle consist in an extension — extrapolation or, sometimes, interpolation — of the observational material, and a comparison with the corresponding values of the functions fitted to the original data.

6. On the linear difference equation.

A class of functions of importance in the sequel, though not as a scheme for time series, is formed by the solutions of linear difference equations with constant coefficients,

$$(x(t)-m) + a_1 \cdot (x(t-1)-m) + \cdots + a_h \cdot (x(t-h)-m) = 0, \quad a_h \neq 0. \tag{32}$$

Writing (30) on the form (32), the resulting sequence a_i will be symmetrical. Now, since (30) is a special case of (32), the solutions of (32) embrace (17), and are well-known to be (see e. g. P. M. MARPLES (1932))

$$(33) \qquad x(t) = m + \sum_{k=1}^{i} H^{(1)}_{m_k-1}(t) \cdot p_k^t + \sum_{k=1}^{j} [H^{(2)}_{n_k-1}(t) \cdot \cos \lambda_k t + \\ + H^{(3)}_{n_k-1}(t) \cdot \sin \lambda_k t] \cdot q_k^t,$$

where $H_r^{(s)}$ stands for a polynomial of order r. While the polynomials H are of arbitrary coefficients, their orders are the same in all solutions. In fact, the orders are determined by the *characteristic equation* of (32), viz.

$$(34) \qquad z^h + a_1 \cdot z^{h-1} + \cdots + a_{h-1} \cdot z + a_h = \\ = \prod_{k=1}^{i} (z - p_k)^{m_k} \cdot \prod_{k=1}^{j} (z^2 + 2 s_k \cdot z + q_k^2)^{n_k} = 0,$$

where the factors in the second member are real.

The frequencies λ_k in (33) are connected with the characteristic equation by the relations

$$(35) \qquad \cos \lambda_k = - s_k / q_k.$$

The asymptotical behaviour of $x(t)$ is dependent on the exponential factors p_k^t and q_k^t, the bases of which are likewise seen to be uniquely determined by (34).

For later application it should be noticed that a necessary and sufficient condition for the convergence of $\sum_{t=1}^{\infty} (x(t) - m)^2$ and $\sum_{t=1}^{\infty} |x(t) - m|$, for any values taken on by the arbitrary coefficients in the polynomials H, is that $|p_k| < 1$ and $|q_k| < 1$ for all k. Another wording of the condition is that all roots of the characteristic equation (34) shall lie within the periphery of the unit circle. In such a case, $x(t)$ will be referred to as describing a *damped oscillation*.

A second property of (32) will also be used later. Let the arbitrary coefficients in (33) be fixed under the single condition that no polynomial H vanishes; then, if some p_k or some q_k is different from unity in modulus, formula (33) shows that $|x(t)|$ cannot possibly be uniformly bounded in $(-\infty < t < \infty)$. In the same way, any solution which does not belong to an equation of lower order is unbounded if (34) presents a multiple root. Thus, if a solution

$x(t)$ of (32) satisfies no linear equation of lower order and if $|x(t)|$ is uniformly bounded in $(-\infty < t < \infty)$, then $|p_k| = |q_k| = m_k = n_k = 1$, i. e. the equation (32) is of the special type (30), and $x(t)$ is of type (17).

7. A purely probabilistic approach.

A typical probabilistic hypothesis about a given time series is to regard the observational values as a random sample of a certain aleatory variable. The complete hypothetical set-up thus consists of a sequence of random variables, say $\ldots, \eta(t-1), \eta(t), \eta(t+1), \ldots$, which are mutually independent, and have identical distribution functions, say $F(x)$.

Since the hypothesis under consideration consists of two elements, the test methods are of two kinds: a) those testing the goodness of fit of the hypothetical function $F(x)$ to the empirical distribution function $\bar{F}(x)$ characterizing the observational series, and b) those testing the randomness in the observational series.

A perfect fit to the data being in contrast to the idea of randomness, an amount of arbitrariness is in place in the choice of an hypothetical distribution function $F(x)$ characterizing the aleatory variables $\eta(t)$.

The classical method for testing goodness of fit is the χ^2-method of K. PEARSON (see e. g. G. U. YULE and M. G. KENDALL (1937), Chapter 22). Another test, viz.

$$\omega^2 = \int_{-\infty}^{\infty} [F(u) - \bar{F}(u)]^2 \, du,$$

has been proposed by H. CRAMÉR ((1927) p. 112, and (1928) p. 145), and, under the term of ω^2-method (Summenlinienverfahren), by R. v. MISES ((1930) p. 316). The latter method, an interesting modification of which has been given by N. SMIRNOFF (1936), does not suffer from the well-known arbitrariness implied in the χ^2-method.[2]

The hypothetical randomness lies in the independence relations of the type (8). Accordingly, the tests of randomness are tests examining various particular instances of these relations. For example, since $r_k = 0$ for $k > 0$, the serial coefficients $\bar{r}_k$ must approximate zero for $k > 0$ (the particular case $k = 1$ is the ABBE-HELMERT criterion — cf. K. STUMPFF (1927), p. 8).

An important instance of the general test problem is concerned

with the choice between functional and probabilistic schemes. This problem is dealt with in the expectance theory (see e. g. K. STUMPFF (1927), p. 115) founded by A. SCHUSTER ((1898), (1900)) and which studies, i. a., the distribution of the periodogram ordinates $C^2(\lambda)$ obtained when substituting a set of independent random variables $\eta(t)$ (cf. p. 11) for (1) in (26). Taking $\bar{m} = E[\eta]$ in (26), the basic formulae of the expectance theory read

$$E[A(\lambda)] = E[B(\lambda)] = 0, \qquad 0 < \lambda \leq \pi. \tag{36}$$

$$E[C^2(\lambda)] = \frac{4}{n} D^2(\xi), \qquad 0 < \lambda < \pi. \tag{37}$$

In case $\lambda = \pi$, the factor 4 in the latter formula must be omitted.

Having now touched upon some typical functional and probabilistic schemes set up for the analysis of time series, we are in a position to pass on to some intermediary schemes. Of these, two types may be distinguished which are of fundamentally different character. Since the terminology does not seem to be established, I propose for the two types the names »schemes of *hidden periodicities*» and »schemes of *linear regression*». The former schemes are the earlier ones, and are dealt with in the next section. Some critical remarks on these schemes follow in section 9. The chapter concludes with some preliminary remarks on the schemes of linear regression.

8. A scheme of hidden periodicities.

A simple approach of hidden periodicities is to regard an observational time series (1) as additively built up by a sum of harmonics and a random sample of a certain aleatory variable. Thus, denoting by

$$\xi(1),\ \xi(2),\ \ldots,\ \xi(n) \tag{38}$$

the hypothetical random variables corresponding to the observational set (1), we have in this case

$$\xi(t) = y(t) + \eta(t), \tag{39}$$

where $y(t)$ is of type (17), and $\eta(t)$ is a set of independent random variables as used in the previous section.

A chief problem when applying a scheme of hidden periodicities is to perform a separation of the functional and the random components. Since we are concerned with the case when $y(t)$ is a composed harmonic (17), the variate difference method, and other methods based on the assumption that $y(t)$ reduces to a polynomial or to a trend function, fall outside of the program (cf. p. 1).

The principal method for the search of harmonic components in a given time series (1) is the SCHUSTER periodogram method described in section 5. The test problem concerning the significance of the ordinates in the empirical periodogram was already mentioned in section 7.

In respect to the OPPENHEIM-BRUNS method for separating harmonic components (cf. section 5), it has been emphasized by J. I. CRAIG (1916) that the method fails when a random component is superposed on the harmonics. This disturbing effect of the random error will be called the »CRAIG *effect*». As the OPPENHEIM-BRUNS method is parallel to a method of importance in the sequel, and as the CRAIG effect — possibly because of the sketchy character of Mr. CRAIG's paper — seems to have been overlooked in later literature, the point in question will be taken up in a separate discussion. This is done in section 28.

The problem of separating the functional and probabilistic elements being, of course, to a considerable extent indeterminate, even rough methods for the search of periodicities may be of interest. A simple method, which is of particular relevance because it is both convenient and capable of delivering more general periodic components than harmonic functions, is based on the well-known BUYS-BALLOT table (see e. g. K. STUMPFF (1927), p. 100):

$$(40)\qquad \begin{cases} \bar{\xi}_1, & \bar{\xi}_2, & \ldots\ldots\ldots\ldots, & \bar{\xi}_p \\ \bar{\xi}_{p+1}, & \bar{\xi}_{p+2}, & \ldots\ldots\ldots\ldots, & \bar{\xi}_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \bar{\xi}_{(k-1)p+1}, & \bar{\xi}_{(k-1)p+2}, & \ldots\ldots\ldots\ldots, & \bar{\xi}_{kp} \\ \bar{\xi}_{kp+1}, & \bar{\xi}_{kp+2}, & \ldots, \bar{\xi}_n & \end{cases}$$

$$(41)\qquad \bar{m}_1, \quad \bar{m}_2, \quad \ldots, \bar{m}_{n-kp}, \ldots, \bar{m}_p.$$

Here p is an integer, k stands for the greatest integer which is less than n/p, and $\bar{m}_i$ is the arithmetical mean of the i:th col-

umn. Denoting by k_i the number of elements in the i:th column, we have

$$\begin{cases} k_i = k + 1 & \text{for } i \leq n - k\,p, \\ k_i = k & \text{for } i > n - k\,p. \end{cases} \tag{42}$$

The leading idea of the method simply is that if the series $\bar{\xi}_t$ contains a component $y(t)$ with period p, the values of $y(t)$ for $t = 1, 2, \ldots, p$ are approximately given by the means $\bar{m}_1, \bar{m}_2, \ldots, \bar{m}_p$.

While the arrangement (40) was used even before C. H. D. Buys-Ballot (1847), the method was developed in detail by B. Stewart and W. Dodgson (1879) and others (cf. H. Burkhardt (1904) p. 679 f.). A sharpening of the method by means of a periodogram construction, generalizing that of A. Schuster, is due to E. T. Whittaker (1911). In the Whittaker periodogram, p is taken for abscissa, and the ordinate in p equals the (weighted) variance in the series $\bar{m}_i$ divided by the variance $\bar{D}^2(\bar{\xi})$ in the series $\bar{\xi}_t$.

The connexion between the two periodogram constructions is interesting. Considering the Schuster periodogram, $\frac{1}{2}\,\bar{C}^2(\lambda)$ is well-known to be approximated by the maximum value for varying A and B of the expression

$$\bar{D}^2(\bar{\xi}) - \frac{1}{n} \cdot \sum_{t=1}^{n} (\bar{\xi}_t - \bar{m} - A \cdot \cos \lambda\, t - B \cdot \sin \lambda\, t)^2. \tag{43}$$

On the other hand, Professor H. Cramér, in his Course in 1933, showed that in the Whittaker periodogram the ordinate equals, apart from the constant denominator $\bar{D}^2(\bar{\xi})$, the maximum of an expression generalizing (43), *viz.*

$$\bar{D}^2(\bar{\xi}) - \frac{1}{n} \cdot \sum_{t=1}^{n} [\bar{\xi}_t - \bar{m} - y_p(t)]^2. \tag{44}$$

This expression should be maximized under the condition that $y_p(t)$ be a function of integral period p, but arbitrary for the rest.*

With the notations used in (40), the maximum value of (44) equals $\frac{1}{n} \cdot \sum_{i=1}^{p} k_i \cdot (\bar{m}_i - \bar{m})^2$, which for the ordinates in the Whittaker periodogram gives

* The interpretation of the generalized periodogram as a correlation ratio is incorrect (cf. E. T. Whittaker and G. Robinson (1926), p. 345 f.).

$$(45) \qquad \sum_{i=1}^{p} k_i (\bar{m}_i - \bar{m})^2 / n \cdot \bar{D}^2(\xi).$$

In case n is a multiple of p, (45) reduces to

$$(45\text{ a}) \qquad \sum_{i=1}^{p} (\bar{m}_i - \bar{m})^2 / p \cdot \bar{D}^2(\xi) = \bar{D}^2(\bar{m}_i) / \bar{D}^2(\xi).$$

In his 1933 Course, Prof. H. Cramér also delivered an expectance theory of the Whittaker periodogram. He showed, i. a., that if the given series is a random sample of a normally distributed variable, then (45 a) is distributed as a variance ratio in Fisher's analysis of variance.

It should be observed that as a principle the Schuster and the Whittaker periodograms are of equal validity — if the former discovers periods in a given observational series, then the latter will give positive results, and *vice versa.*

9. On the criticism of the scheme of hidden periodicities.

While the hypothesis of hidden periodicities has proved very fruitful in many fields of scientific research, many applications early met with a serious criticism (cf. H. Burkhardt (1904) p. 685). The essential point of criticism bears upon the postulated strict periodicity of the individual functional components, and it has been maintained that this rigidity in the periodicity often has no empirical correspondence.

The serial and autocorrelation coefficients disclose an interesting aspect of the above critical argument. If the functional element $y(t)$ in (39) is a composed harmonic (17), the autocorrelation coefficients r_k are for $k \neq 0$ given by

$$(46) \qquad r_k = \sum_{i=1}^{s} C_i^2 \cdot \cos \lambda_i k / (2\, D^2(\eta) + \sum_{1}^{s} C_i^2).$$

The hypothesis of hidden periodicities assumes, therefore, that, for $k \neq 0$, the autocorrelation coefficient r_k, too, is a function of the type (17), i. e, a composed harmonic. According to the relation (19), the hypothesis thus implies that there exist arbitrarily large k-values such that r_k is approximately given by the value taken by (46) for $k = 0$, viz.

$$(47) \qquad \sum_1^s C_i^2 / (2\, D^2(\eta) + \sum_1^s C_i^2).$$

This implication may, as a matter of fact, be used as a criterion of the applicability of the hypothesis of hidden periodicities. Thus, even though a given time series clearly shows a cyclical character, but the serial coefficients are gradually vanishing, then the scheme of hidden periodicities is no adequate approach.

It seems plausible that in a good many oscillatory phenomena the serial coefficients actually are gradually vanishing. The above criterion shows that in such cases a periodogram analysis would give negative results. The table of serial coefficients in air pressure material from *Port Darwin* analyzed by SIR G. WALKER ((1931), p. 528) may be referred to for illustration. The graph of serial coefficients presents damped oscillations of a period of about three years.

It follows from the above that in a descriptive analysis of a time series the serial coefficients of G. U. YULE are of fundamental importance. J. BARTELS (1935) has recently given another method of descriptive analysis, consisting in a generalization of the BUYS-BALLOT table, and directly constructed as a criterion of the applicability of the hypothesis of hidden periodicities. J. BARTELS forms the BUYS-BALLOT table for successively extended sections of the given series. Let $\theta(k, \nu)$ stand for the ν:th section consisting of k consecutive rows in (40), and let q be the number of such sections. Let $\bar{D}_p^2(k, \nu)$ be the variance of the column averages $\bar{m}_i(k, \nu, p)$ in the section $\theta(k, \nu)$, and let their arithmetical mean in respect of ν be $\bar{J}_p(k)$,

$$\bar{J}_p(k) = \frac{1}{q} \cdot \sum_{\nu=1}^{q} \bar{D}_p^2(k, \nu).$$

For the expression

$$(48) \qquad \bar{B}_p(k) = k \cdot \bar{J}_p(k) / \bar{J}_p(1),$$

regarded as a function of k, J. BARTELS proposes the term *persistency characteristic* because of the following observations.

Let the series ξ_t under investigation contain a component of persistent period, i.e. a strictly periodic component. If the length of the period equals p, and if k is large, each of the q sets $\bar{m}_i(k, \nu, p)$, considering the variation with i, will nearly reproduce the periodic

component. Consequently, $\bar{J}_p(k)$ will tend to a positive constant as $k \to \infty$, and $\bar{B}_p(k)$ will therefore increase nearly proportionately with k. — On the other hand, if the time series is purely random, $\bar{B}_p(k)$ will show no tendency to vary with k, but will remain on the unity level. — Concerning intermediate cases, J. BARTELS remarks that $\bar{B}_p(k)$ may tend to an asymptote above unity. Then a quasi-persistency is present, a tendency of adjacent rows in the BUYS-BALLOT table to show a certain resemblance, a resemblance which will fade away as more distant rows are compared.

The persistency characteristics computed by J. BARTELS ((1935) p. 519 f.) suggest persistency in statistical data concerning a) the half year period in the international index of terrestrial magnetism, b) the 24 hour wave in air pressure in *Potsdam*, c) the 12 hour component of the *Batavia* temperature. On the other hand, quasi-persistency is suggested in a) the 27 day component in terrestrial magnetism, b) the 24 hour wave in the magnetic east component in *Batavia*.

In economics, the classical periodogram analysis has repeatedly been tried on business cycle material. The negative results support an opinion which has been maintained also on logical-theoretical grounds, and which now seems predominant, *viz.* that the hypothesis of hidden periodicities is inadequate in business cycle theory.

In cases like those mentioned above, where the approach of rigid periodicity fails, the schemes of linear regression announced in section 7 seem to form a natural and interesting substitute for the scheme of hidden periodicities. Reference to previous results concerning the schemes of linear regression being given later, when dealing systematically with these schemes, the concluding section in this survey will enter into detail only as to the earliest papers on the schemes in question.

10. Remarks on the schemes of linear regression.

In the hypothesis of hidden periodicities, there is assumed a fargoing interdependence between the elements of the given time series; leaving the random component out of the question, the interdependence is assumed to be purely functional. The schemes of linear regression assume as to adjacent elements an interdependence only in the sense of the theory of probability.

In an interesting study on the variate difference method, G. U. YULE (1921) considers the autocorrelation in a series consisting of iterated differences, say of order m, obtained from a purely random series. Since the autocorrelation coefficients are

$$(49) \qquad r_k = (-1)^k \cdot \frac{m(m-1)\cdots(m-k+1)}{(m+1)\cdot\cdots\cdot(m+k)}; \qquad 0 < k \leq m,$$

the series of differences must present an oscillatory character, a feature increasing in evidence with m. In other words, we are concerned with a primary sequence of random variables, say ..., $\eta(t-1)$, $\eta(t)$, $\eta(t+1)$, ..., which by hypothesis are independent, and have identical distribution functions; on this basis a secondary series, say ..., $\xi(t-1)$, $\xi(t)$, $\xi(t+1)$, ..., is defined by means of a moving linear operation of the type

$$(50) \qquad \xi(t) = b_0\,\eta(t) + b_1\,\eta(t-1) + \cdots + b_h\,\eta(t-h).$$

The approach thus defined will in the sequel be called the scheme of *moving averages*.

Another particular case of moving average (50) is studied by E. SLUTSKY (1927), who forms the secondary, intercorrelated series from the primary, purely random series by n iterated summations by two, followed by the forming of m:th differences. Holding m/n constant, E. SLUTSKY shows that, with probability 1, an arbitrarily fixed section of the difference series will tend to a sine curve as $n \to \infty$. This result is given as an application of a general theorem proved in the same paper and discussed in section 16.

Stochastical interdependence of the type (50) is a particular case of linear regression. Letting the auxiliary variables $\eta(t)$ be the same, another type of linear regression is indicated by the following implicit definition of the intercorrelated variables $\xi(t)$,

$$(51) \qquad \xi(t) + a_1\,\xi(t-1) + \cdots + a_h\,\xi(t-h) = \eta(t).$$

The approach (51) was introduced in an heuristic manner in an important memoir by G. U. YULE (1927). The fundamental difference between the scheme of hidden periodicities and the scheme (51), which is said by YULE to define a »disturbed harmonic» $\xi(t)$, is clearly brought out. In (39) the random elements $\eta(t)$ *»do not in any way disturb the steady course of the underlying periodic function or functions»* (p. 268). On the other hand, regarding (51) YULE

(p. 294) states that a principal feature of a disturbed periodic movement is »*a continual change of amplitude and shift of phase*».

In the same paper G. U. YULE applies, with success, his hypothesis to empirical data, *viz.* A. WOLFER's sunspot numbers. Further, he makes the following general statement concerning the scope of the scheme (51): »*Disturbance will always arise if the value of the variable is affected by external circumstance and the oscillatory variation with time is wholly or partly self-determined, owing to the value of the variable at any one time being a function of the immediately preceding values. Disturbance, as it seems to me, can only be excluded if either (1) the variable is quite unaffected by external circumstance, or (2) we are dealing with a forced vibration and the external circumstances producing this forced vibration are themselves undisturbed*» (p. 295).

In order to attain conformity in terminology, the approach (51) will in the sequel be descriptively called the scheme of *(linear) autoregression*. G. U. YULE (1927) restricts himself to the cases $h \leq 4$. General autoregression as implicitly defined by (51) was dealt with by SIR G. WALKER (1931), whose applications to the air pressure data mentioned in section 9 gave positive results.

As shown in detail by E. SLUTSKY (see e.g. (1937)), even a scheme of moving averages (50) may present waves of shifting amplitude and phase. Thus, both schemes of linear regression are of interest to the theory of those oscillatory phenomena for which the hypothesis of hidden periodicities proves inadequate. The investigations referred to in section 9 show that there are many central phenomena of this kind.

While the schemes of linear regression thus form a type of hypothesis of the greatest importance, the development of the subject is still little advanced, both as to the theory and the application of the schemes. For instance, earlier definitions concerning the scheme of autoregression are incomplete. One of the chief purposes of the present volume is to give some contributions for completion in these respects. It also aims at bringing the schemes into place in the theory of probability, thereby uniting the rather isolated results hitherto reached.

In the theory of probability, the schemes of linear regression fall under the heading of the discrete stationary random process as defined by A. KHINTCHINE (see (1932) and (1933)). As a matter of fact, the concept of stationary random process is extremely general, and the restrictions involved are only those indispensable

in hypotheses concerning stationary phenomena (cf. p. 3). It is therefore but natural that the scheme of hidden periodicities, after a slight change in the interpretation, will also be found to form a special discrete stationary process (see section 15 Ω). Accordingly, the theoretical developments start with a chapter on the discrete random process. This analysis will, i.a., deepen the insight into the nature of the schemes of linear regression, which are dealt with in Chapter III. Chapter IV contains some applications of different hypotheses to observational time series.

The present work is confined to the time-series aspects of the empirical data. Since the serial coefficients play a fundamental part in the analysis we note by way of a general caution that, as always when dealing with correlation coefficients, it is not a matter of pure routine to apply the general methods in practice, and the field is full of pitfalls. For one thing, there is the question of the quantitative significance of correlation coefficients. Following up an argument presented in a preliminary note [see H. Wold (1936)], this question was taken up for discussion in Appendix A of the 1st edition of this book. To restate the main conclusion, this is that for correlation coefficients which are obtained from a broad class of time or spatial series their quantitative significance is influenced by the size of the statistical masses to which the data refer. In particular, this conclusion applies to a broad class of serial coefficients.

CHAPTER II.

On the theory of the discrete stationary random process.

11. Definition of the stationary processes.

In the theoretical analysis of a statistical time series, we may distinguish between functional and probabilistic approaches. In the former, the time series is represented by a function of time which in the general case is univalent, but otherwise unconditioned. In the latter, the most general approach is the unconditioned random process. From a purely mathematical viewpoint, the random process is a random variable in an infinite number of dimensions. Denoting by $\{t\}$ the set of time points t in which the phenomenon changing with time is studied, each element in $\{t\}$ corresponds to one dimension in the random variable.

Let $\{t\}$ stand for a set of values taken on by a real parameter t, which will be spoken of as representing time, and let one random variable $\xi(t)$ correspond to each time point t in $\{t\}$. Denoting the given set of random variables by $\{\xi(t)\}$, let it be assumed that the following conditions are satisfied.

(A). Choosing arbitrarily a sub-set in $\{t\}$, say $(t) = (t_1, \ldots, t_n)$, the combined variable $\xi(t_1, \ldots, t_n) = [\xi(t_1), \ldots, \xi(t_n)]$ will be well-defined.

Let the distribution function of $\xi(t_1, \ldots, t_n)$ be denoted by $F(t_1, \ldots, t_n; u_1, \ldots, u_n)$, so that

$$(52) \qquad F(t_1, \ldots, t_n;\ u_1, \ldots, u_n) = P[\xi(t_1) \leq u_1, \xi(t_2) \leq u_2, \ldots, \xi(t_n) \leq u_n],$$

and let the sets of distribution functions and probability functions of the variables $\xi(t_1, \ldots, t_n)$ be denoted by $\{F\}$ and $\{P\}$ respectively.

(B). Letting $(t) = (t_1, \ldots, t_n)$ be an arbitrary sub-set in $\{t\}$, and $(i_1, \ldots, i_n)$ be an arbitrary permutation of the sequence $(1, 2, \ldots, n)$, the functions $\{F\}$ will satisfy the following relations identically in $u_1, \ldots, u_n$,

$$(53)\quad F(t_{i_1}, t_{i_2}, .., t_{i_n};\quad u_{i_1}, u_{i_2}, .., u_{i_n}) = F(t_1, .., t_n;\quad u_1, .., u_n),$$

$$(54)\quad F(t_1, .., t_m;\quad u_1, .., u_m) = F(t_1, .., t_n;\quad u_1, .., u_m, +\infty, .., +\infty),$$

where $m < n$.

These relations express merely that the probability laws ruling the variables $\{\xi(t)\}$ must not contradict themselves. Accordingly, these relations will be referred to as the *consistency relations*.

Following A. Kolmogoroff ((1931), (1933)), a set $\{\xi(t)\}$ satisfying the conditions (A) and (B) will be called a *random process*.

According to a fundamental theorem of A. Kolmogoroff ((1933), p. 27), a set $\{F\}$ belonging to a random process $\{\xi(t)\}$ defines a probability distribution on those sets in a space R_t of an infinite number of dimensions $\{t\}$, which are formed by an enumerable sum of Borel's cylinder sets in R_t. For instance, if the sequence $t, t-1, t-2, \ldots$ is contained in the set $\{t\}$, the probability $P[\xi(t) \leq u_0, \xi(t-1) \leq u_1, \xi(t-2) \leq u_2, \ldots]$ will exist for any real sequence $u_0, u_1, u_2, \ldots$

We see that the stochastic process extends the notion of random variable from a finite to an infinite number of dimensions. The frequency interpretation in terms of a universe of sample elements remains the same. The sample elements of a process $\{\xi(t)\}$, also called *realizations* of the process, are functions of t, say $\xi_i(t)$. Considering the universe of realizations, and keeping t fixed, say $t = t_1$, we obtain the universe of sample values $\xi_i(t_1)$ that constitute the random variable $\xi(t_1)$. More generally, if we keep $t_1, \ldots, t_n$ fixed, the realizations will give us the universe of sample elements $[\xi_i(t_1), \ldots, \xi_i(t_n)]$ that constitute the n-dimensional random variable $[\xi(t_1), \ldots, \xi(t_n)]$.

In order to define stationarity, we must consider arbitrary translations within the set $\{t\}$. In doing this, we can assume that $\{t\}$ either consists of all real t-values, or is formed by an unbroken sequence of equidistant values, say $\ldots, -1, 0, 1, 2, \ldots$ In any case, a random process $\{\xi(t)\}$ as defined by a set $\{F\}$ is termed *stationary* in the sense of A. Khintchine ((1932)—(1934)), if for an arbitrary sub-set $(t) = (t_1, .., t_n)$ in $\{t\}$ the relation

$$(55)\quad F(t_1 + t, t_2 + t, \ldots, t_n + t;\quad u_1, .., u_n) = F(t_1, t_2, \ldots, t_n;\quad u_1, .., u_n)$$

is identically satisfied in $u_1, .., u_n$ and in t. Again following Khintchine, the process will be called *discrete*, if t is restricted to a sequence of equidistant values, *continuous* if t is arbitrary.

According to the interpretation indicated in section 2, the variables $\xi(t)$ considered in the above definitions may be taken to be multidimensional. Thus, just as in the case of ordinary random variables, a k-dimensional random process $\{\xi(t)\}$ may be looked upon as obtained by combining k one-dimensional processes, say $\{\xi^{(1)}(t)\}, \ldots, \{\xi^{(k)}(t)\}$. Now, in studying simultaneously a group of one-dimensional processes, we shall always assume that an arbitrary finite sub-group, say $\{\xi^{(i_1)}(t)\}, \ldots, \{\xi^{(i_k)}(t)\}$, can be combined into a k-dimensional random process. For instance, considering an infinite sequence $\{\xi^{(1)}(t)\}, \{\xi^{(2)}(t)\}, \ldots$, this assumption may be expressed as follows. Taking out arbitrarily a group $\{\xi^{(i_1)}(t)\}, \ldots, \{\xi^{(i_k)}(t)\}$, fixing arbitrarily a set of time points $t_1, \ldots, t_n$, and a double real sequence $u_s^{(r)}$, where $r = 1, 2, \ldots, k$; $s = 1, 2, \ldots, n$, we shall assume that the probability

$$P[\xi^{(i_r)}(t_s) \leq u_s^{(r)}; \quad r = 1, \ldots, k; \quad s = 1, \ldots, n]$$

will exist; further, we shall assume that these probabilities will satisfy all consistency relations of type (53—54); in case of stationarity we shall also assume that all relations of type (55) will hold.

Expectations derived from the distribution functions $\{F\}$ determining a stationary process $\{\xi(t)\}$ will be called *characteristics* of the process. The characteristics are, of course, independent of t. Further, considering the distribution functions $F(t;\, u)$ in the set $\{F\}$, these will be independent of t. The function of u thus obtained will be termed the *principal* distribution function of the process considered. By definition, the *mean* (m), the *dispersion* (D), etc., of a one-dimensional stationary process are given by the corresponding characteristics as obtained from the principal distribution function (cf. (5) and (6)).

If the dispersion of a one-dimensional stationary process is finite, the *automoments* of second order as defined by (cf. (10) and (11))

$$\nu_2^{(k)} = E[\xi(t) \cdot \xi(t+k)] = \int_{R_2} u\, v.\, d_{u,v} F(t,\, t+k;\ u,\, v) = \nu_2^{(-k)}$$

will be finite. The characteristics mentioned determine the *autocorrelation coefficients* belonging to the stationary process $\{\xi(t)\}$ considered (cf. (12)),

$$(56) \qquad r_k = r_k(\xi) = (\nu_2^{(k)} - m^2)/D^2 = r_{-k}$$

If $r_k(\xi) = 0$ for all $k \neq 0$, the process $\{\xi(t)\}$ will be termed *non-autocorrelated.*

A. KHINTCHINE (1932) gives also a more embracing definition of the stationary process, requiring only that the characteristics m, D and $\nu_2^{(k)}$ shall be independent of t. This case will be referred to as the *generalized* stationary process.

Let $\{\xi(t)\}$ be a stationary process, and consider the variables $\xi(t_1, \ldots, t_m)$ and $\xi(t_1, \ldots, t_n)$ which refer to the time points $(t_1, \ldots, t_m, \ldots, t_n)$. We shall sometimes have to regard the process $\{\xi(t)\}$ in the set $(t_{m+1}, \ldots, t_n)$ as conditioned by the behaviour of $\{\xi(t)\}$ in the time points $(t_1, \ldots, t_m)$. Following a familiar terminology, we shall then speak of the variable $\xi(t_{m+1}, \ldots, t_n)$ as being *conditioned* by $\xi(t_1, \ldots, t_m)$. Indicating conditionality by an index C, and denoting by C' the condition obtained from C by replacing throughout t_i by $t_i + t$, it is evident that for an arbitrarily fixed t the two conditioned variables $\xi_C(t_{m+1}, \ldots, t_n)$ and $\xi_{C'}(t_{m+1} + t, \ldots, t_n + t)$ will have identical distribution functions if the process $\{\xi(t)\}$ is stationary. The reader is referred to A. KOLMOGOROFF ((1933), Chapter V) for the fundamentals concerning conditioned variables and distributions.

Generally speaking, a characteristic of a conditioned variable depends on the conditioning variable, and will be called a *conditioned characteristic*. Since such a characteristic forms a function of a random variable, it constitutes in itself a random variable. For instance, considering a one-dimensional process $\{\xi(t)\}$, and taking as before $\xi(t_1, \ldots, t_m)$ to be the conditioning variable, the conditioned expectation of $\xi(t_{m+1})$ is, by definition, the expectation of $\xi_C(t_{m+1})$. Denoting this expectation by $E_C[\xi(t_{m+1})]$, a general formula gives (see A. KOLMOGOROFF (1933), p. 47)

$$E[E_C[\xi(t_{m+1})]] = E[\xi(t_{m+1})] = m(\xi). \tag{57}$$

From now on, when not explicitly stated otherwise, the random processes dealt with are tacitly understood (A) *to be one-dimensional,* (B) *to be discrete, and defined for integral time points,* (C) *to have a finite dispersion.* Of course, the confinement to integral time points instead of a general equidistant sequence, say

$$\ldots, t_0 - a, \ t_0, \ t_0 + a, \ t_0 + 2a, \ldots$$

involves no restriction as to the generality of the theory of the discrete process.

12. A theorem of A. Khintchine.

As far as I am aware, the only earlier investigation of the general discrete stationary process is that of A. Khintchine ((1932), (1933)) already referred to. Though the problems dealt with in the sequel lie along entirely different lines, the principal theorem of Khintchine will be quoted in full because it discloses a fundamental property of the stationary process. The theorem in question states that the stationary process is subjected to the law of great numbers, viz. in the following sense:

Let $\xi(t)$, $\xi(t-1)$, ..., $\xi(t-n+1)$ *be a finite sequence of variables connected with a discrete stationary process* $\{\xi(t)\}$ *with finite dispersion, and put* $\Sigma_n = \frac{1}{n} \cdot \sum_{i=0}^{n-1} \xi(t-i)$. *The dispersion of the difference* $\Sigma_n - \Sigma_m$ *then tends to zero when* $n \to \infty$ *and* $m \to \infty$. *Further, processes may be constructed so that the asymptotical decrease is arbitrarily slow.*[3]

In view of the applications — in particular certain questions concerning ergodic hypotheses — it is an interesting problem whether in the sums Σ the sequence $i = 0, 1, 2, \ldots$ can be replaced by the sequence $i_0 < i_1 < i_2 < \cdots$ A short reflection on the singular processes as defined and exemplified in sections 14—16 shows that such a general sequence is not allowed.

The problems dealt with in the sequel have their points of connection with certain investigations on the continuous process and on special types of the discrete process. These earlier investigations will be referred to in the course of the analysis.

The concept of stationary process as introduced by A. Khintchine is extremely general. As a scheme for the analysis of time series it will be found to embrace all the schemes mentioned in the survey given in Chapter I. For the sake of concreteness it will be of interest, before passing to the general theoretical developments, to show in detail how these may be obtained by suitable specializations. To this end a preparatory analysis of the general process will be useful. Accordingly, the next two sections will be reserved for some groundwork concerning operation with the discrete stationary processes.

13. Some fundamental operations with random processes.

In this section it will be shown that certain familiar operations with ordinary random variables can also be performed with random

processes. In discussing the situation, no restrictions will be laid on the processes considered. Generally speaking, the operations will give rise to new random processes. Moreover, if the processes dealt with are stationary, the resulting processes will be found to be stationary. It will be sufficient for our purpose to consider functions of random processes, and the forming of limit processes in convergent sequences.

Denoting by ξ a random variable in a k-dimensional space R_k, and by $f[x]$ a function which is finite and BOREL-measurable in R_k, and whose values are lying in a space R_p, it is known that $f[\xi]$ will be a well-defined random variable in R_p (see e. g. H. CRAMÉR (1937), p. 12 f). Let us next consider a combined variable $[\xi^{(1)}, \ldots, \xi^{(n)}]$ consisting of random variables $\xi^{(i)}$ in R_k. Forming the variables $f[\xi^{(i)}]$, it is evident that these in the same way may be combined to a random variable $[f[\xi^{(1)}], \ldots, f[\xi^{(n)}]]$ (cf. also p. 10).

Thus prepared, let $f[x]$ remain the same, and consider the variables $\xi(t_1, \ldots, t_n) = [\xi(t_1), \ldots, \xi(t_n)]$ which constitute a random process $\{\xi(t)\}$. According to the above, the variables $[f[\xi(t_1)], \ldots, f[\xi(t_n)]]$ will be well-defined. Denoting by $\{F^*\}$ the set of distribution functions of these variables, it is also evident that the functions F^* satisfy all consistency relations of type (53—54). Further, if $\{\xi(t)\}$ is stationary, the functions F^* also will satisfy (55). The variables $[f[\xi(t_1)], \ldots, f[\xi(t_n)]]$ will thus constitute a random process, and if $\{\xi(t)\}$ is stationary, the process obtained will also be stationary. The resulting process will be said to be a function of the process $\{\xi(t)\}$, and be denoted by $\{f[\xi(t)]\}$. The variables of type $[f[\xi(t)], f[\xi(t-1)], \ldots, f[\xi(t-n)]]$ will be denoted by $f[\xi(t, t-1, \ldots, t-n)]$.

In particular, considering a random process $\{\xi(t)\}$ obtained by combining k one-dimensional processes, say $\{\xi^{(1)}(t)\}, \ldots, \{\xi^{(k)}(t)\}$, let us take f to be linear. The operation will then give rise to a sum process of type $\{a_1 \xi^{(1)}(t) + \cdots + a_k \xi^{(k)}(t)\}$. Denoting this sum by $\{\zeta_k(t)\}$, we shall write

$$\{\zeta_k(t)\} = a_1 \{\xi^{(1)}(t)\} + \cdots + a_k \{\xi^{(k)}(t)\}.$$

According to the above, we have

$$(58) \quad \zeta_k(t_1, \ldots, t_n) = [a_1 \xi^{(1)}(t_1) + \cdots + a_k \xi^{(k)}(t_1), \ldots, a_1 \xi^{(1)}(t_n) + \cdots + a_k \xi^{(k)}(t_n)] =$$
$$= a_1 \xi^{(1)}(t_1, \ldots, t_n) + \cdots + a_k \xi^{(k)}(t_1, \ldots, t_n).$$

Most of the functional operations dealt with in the sequel are of this simple kind. As to the processes $\{\xi^{(1)}(t)\}, \ldots, \{\xi^{(k)}(t)\}$ combined, these, too, are for the most part of a simple structure. We shall next present two type cases.

Let $\{\xi^{(1)}(t)\}, \ldots, \{\xi^{(k)}(t)\}$ represent a set of random processes and let a set of time points $(t) = (t_1, \ldots, t_n)$ and k sets of real numbers $(u^{(s)}) = (u_1^{(s)}, \ldots, u_n^{(s)})$; $s = 1, .., k$, be chosen arbitrarily. Then the processes $\{\xi^{(i)}\}$ will be called *independent* if the following relation is satisfied

$$P[\xi^{(1)}(t_1) \le u_1^{(1)}, .., \xi^{(1)}(t_n) \le u_n^{(1)}; \ \ldots; \ \xi^{(k)}(t_1) \le u_1^{(k)}, .., \xi^{(k)}(t_n) \le u_n^{(k)}] =$$

$$= P[\xi^{(1)}(t_1) \le u_1^{(1)}, .., \xi^{(1)}(t_n) \le u_n^{(1)}] \ldots . P[\xi^{(k)}(t_1) \le u_1^{(k)}, .., \xi^{(k)}(t_n) \le u_n^{(k)}].$$

Similarly, a sequence $\{\xi^{(1)}(t)\}, \{\xi^{(2)}(t)\}, \ldots .$ will be said to consist of independent processes if the processes in every finite subset $\{\xi^{(i_1)}(t)\}, \ldots, \{\xi^{(i_k)}(t)\}$ are independent.

Now, let it be assumed that the independent processes $\{\xi^{(i)}(t)\}$ are stationary, and have finite dispersions $D(\xi^{(i)})$. Then the sum process $\{\zeta_k(t)\}$ as defined by (58) will be stationary, and the expectation, the dispersion, and the autocorrelation coefficients of $\{\zeta_k(t)\}$ will exist. We have $E[\zeta_k] = a_1 E[\xi^{(1)}] + a_2 E[\xi^{(2)}] + \cdots + a_k E[\xi^{(k)}]$, and, as is readily verified,

$$(59) \qquad \begin{cases} D^2(\zeta_k) = a_1^2 D^2(\xi^{(1)}) + a_2^2 D^2(\xi^{(2)}) + \cdots + a_k^2 D^2(\xi^{(k)}), \\ r_p(\zeta_k) = a_1^2 \dfrac{D^2(\xi^{(1)})}{D^2(\zeta_k)} \cdot r_p(\xi^{(1)}) + \cdots + a_k^2 \dfrac{D^2(\xi^{(k)})}{D^2(\zeta_k)} \cdot r_p(\xi^{(k)}). \end{cases}$$

Of course, the two latter relations depend on the identities

$$(60) \qquad r(\xi^{(r)}(t \pm p); \ \xi^{(s)}(t \pm q)) = 0, \qquad p \ge 0, \ q \ge 0,$$

where r and s are arbitrary. Having stated this, $\{\xi^{(r)}\}$ and $\{\xi^{(s)}\}$ will be termed *non-correlated* or *uncorrelated* if the relations (60) are satisfied. Evidently, the relations (59) hold under the broader assumption that any two processes $\{\xi^{(r)}\}$ and $\{\xi^{(s)}\}$ are non-correlated.

In order to define the second type case, let us consider the variables $\xi(t_1, .., t_n)$ which constitute a random process $\{\xi(t)\}$. Choosing arbitrarily an integer k, let us form a second type of variable, say $\xi^*(t_1, .., t_n)$, by taking $\xi^*(t_1, .., t_n) = \xi(t_1 + k, .., t_n + k)$. Evidently,

these variables ξ^* constitute a random process. Now, denoting this process by $\{\xi(t+k)\}$, a short reflection shows that we may combine the processes $\{\xi(t)\}$ and $\{\xi(t+k)\}$. In fact, the distribution functions ruling the simultaneous behaviour of $\{\xi(t)\}$ and $\{\xi(t+k)\}$ are uniquely determined by the distribution functions $\{F\}$ ruling the process $\{\xi(t)\}$, and it is further evident that the resulting distribution functions will satisfy all consistency relations (53—54). Moreover, if $\{\xi(t)\}$ is stationary, the combined process will be stationary.

The arguments being perfectly general, we can form the processes $\{\xi(t)\}$, $\{\xi(t-1)\}$, ..., $\{\xi(t-h)\}$ and combine them into an $(h+1)$-dimensional process. Now, if $\{\xi(t)\}$ is one-dimensional, we can apply a linear operation of type (58) to the combined process. The result will be a process, say $\{\zeta(t)\}$, such that the corresponding variables $\zeta(t)$ and $\zeta(t_1, \ldots, t_n)$ will satisfy relations like

$$\zeta(t) = a_0 \xi(t) + a_1 \xi(t-1) + \cdots + a_h \xi(t-h) \tag{61}$$

$$\zeta(t_1, \ldots, t_n) = [a_0 \xi(t_1) + a_1 \xi(t_1 - 1) + \cdots + a_h \xi(t_1 - h), \\ \ldots, a_0 \xi(t_n) + a_1 \xi(t_n - 1) + \cdots + a_h \xi(t_n - h)]. \tag{62}$$

If $\{\xi(t)\}$ is stationary, then $\{\zeta(t)\}$ will also be stationary.[4]

The operations considered above may also be applied to observational time series. Letting ..., $\bar{\xi}_{t-1}, \bar{\xi}_t, \bar{\xi}_{t+1}, \ldots$ represent such a series, and transforming by means of a function $f[x]$, the resulting series will read ..., $f[\bar{\xi}_{t-1}], f[\bar{\xi}_t], f[\bar{\xi}_{t+1}], \ldots$. If, in particular, every $\bar{\xi}_t$ consists of a couple of k observations, say $(\bar{\xi}_t^{(1)}, \ldots \bar{\xi}_t^{(k)})$, and if $f[x]$ is linear, the transformed series, say ..., $\bar{\zeta}_{t-1}, \bar{\zeta}_t, \bar{\zeta}_{t+1}, \ldots$, will have for general element

$$\bar{\zeta}_t = a_1 \bar{\xi}_t^{(1)} + a_2 \bar{\xi}_t^{(2)} + \cdots + a_k \bar{\xi}_t^{(k)}.$$

On the other hand, assuming $\bar{\xi}_t$ to be one-dimensional, the transform (61) corresponds simply to a moving linear operation. In this case the general element in the transformed series reads

$$\bar{\zeta}_t = a_0 \bar{\xi}_t + a_1 \bar{\xi}_{t-1} + \cdots + a_h \bar{\xi}_{t-h}. \tag{63}$$

In the theoretical developments, we shall frequently have to consider sums of type (58) when the number of terms tends to infinity. For use in such connexions, we need a suitable definition of convergence. To this end we shall next extend the concept of con-

vergence in probability, as introduced by F. P. CANTELLI (1916), so as to apply to a sequence of random processes.

A sequence $\xi^{(1)}, \xi^{(2)}, \ldots$ of ordinary random variables is said to *converge in probability* to a random variable ξ if for every $\varepsilon > 0$

$$P[|\xi^{(n)} - \xi| > \varepsilon]$$

tends to zero as $n \to \infty$. A necessary and sufficient condition for such convergence is that, for every $\varepsilon > 0$, there exists a number n such that for an arbitrary $q > 0$ the following inequality holds (see A. KOLMOGOROFF (1933), p. 32).

$$P[|\xi^{(n+q)} - \xi^{(n)}| > \varepsilon] < \varepsilon.$$

Using the familiar interpretation of $|x - y|$ as the distance between two points x and y in a multi-dimensional space, the definition of convergence also holds in case the variables are k-dimensional, say $\xi^{(i)} = [\xi_1^{(i)}, \ldots, \xi_k^{(i)}]$, $\xi = [\xi_1, \ldots, \xi_k]$. Of course, an equivalent definition is that for every $\varepsilon > 0$ the probability

$$P[|\xi_1^{(n)} - \xi_1| < \varepsilon, \ldots, |\xi_k^{(n)} - \xi_k| < \varepsilon]$$

tends to unity as $n \to \infty$. Now, considering the elementary inequality of G. BOOLE,

$$\begin{aligned} (64) \quad P[|\xi_r^{(n)} - \xi_r| > \varepsilon] \leq 1 - P[|\xi_1^{(n)} - \xi_1| \leq \varepsilon, \ldots, |\xi_k^{(n)} - \xi_k| \leq \varepsilon] \leq \\ \leq P[|\xi_1^{(n)} - \xi_1| > \varepsilon] + \cdots + P[|\xi_k^{(n)} - \xi_k| > \varepsilon], \end{aligned}$$

it is evident that a necessary and sufficient condition that $\xi^{(n)}$ converges in probability to ξ is that $\xi_r^{(n)}$ converges in probability to ξ_r for $r = 1, \ldots, k$ (see F. P. CANTELLI (1916) and E. SLUTSKY (1925)).

Thus prepared, let a sequence of random processes be denoted by

$$(65) \qquad \{\xi^{(1)}(t)\}, \ \{\xi^{(2)}(t)\}, \ldots.$$

The sequence will be called *convergent in probability* to a *limit process* $\{\xi(t)\}$ if for an arbitrary set $(t) = (t_1, \ldots, t_n)$ the sequence

$$(66) \qquad \xi^{(1)}(t_1, \ldots, t_n), \ \xi^{(2)}(t_1, \ldots, t_n), \ldots.$$

be convergent in probability to the limit variable $\xi(t_1, \ldots, t_n)$.

Theorem 1. A necessary and sufficient condition that a sequence (65) *of random processes be convergent in probability is that for an arbitrary t the sequence*

$$\xi^{(1)}(t),\ \xi^{(2)}(t),\ \ldots . \tag{67}$$

be convergent in probability. If the sequence (65) *is convergent, and if every process* $\{\xi^{(n)}(t)\}$ *is stationary, the limit process will be stationary.*

The necessity of the condition is implied in the above definition of convergence in probability. Next, let $(t) = (t_1, \ldots, t_n)$ be arbitrarily fixed, and consider the sequence (66). According to the above application of the inequality of G. Boole, the convergence of (66) is implied in the convergence of (67) for every t in the set $(t_1, \ldots, t_n)$. Having stated this, let the limit variable be denoted by $\xi(t_1, \ldots, t_n)$, and consider the limit variables belonging to all possible sets $(t) = (t_1, \ldots, t_n)$. Since the consistency relations of type (53—54) are satisfied for every process in the sequence (65), the same relations must be satisfied in the limit. The variables $\xi(t_1, \ldots, t_n)$ will thus constitute a random process, say $\{\xi(t)\}$, and the same argument shows that the limit process $\{\xi(t)\}$ is stationary if every process in the sequence (65) is stationary.

The following corollary needs no comment.

Corollary. Letting $\{\xi(t)\}$ *represent a random process, the sequence* (65) *converges in probability to* $\{\xi(t)\}$ *if, and only if, for an arbitrary t the sequence* (67) *converges in probability to* $\xi(t)$.

In dealing with stationary sequences (65), the theorem proved above will be particularly useful, for the behaviour of (67) in respect of convergence will then be independent of t. Considering, in particular, the sum process $\{\zeta_k(t)\}$ defined by (58), a necessary and sufficient condition for convergence in probability as $k \to \infty$ is that the sum $\sum\limits_{i=1}^{\infty} a_i \xi^{(i)}(t)$ is convergent. Reference is made to A. Kolmogoroff (1933) for the groundwork on infinite series of random variables.

A sufficient condition of convergence is, of course, that the sum $\sum\limits_{1}^{\infty} a_i \cdot E[\xi^{(i)}(t)]$ is convergent, and that the dispersion of $\sum\limits_{i=n}^{n+p} a_i \xi^{(i)}(t)$

tends to zero uniformly in p as $n \to \infty$. Moreover, writing m_n and σ_n for the mean and dispersion of $\sum_{i=1}^{n} a_i \cdot \xi^{(i)}(t)$, it follows readily that if these conditions are satisfied, it would imply a contradiction were not the mean and dispersion of the limit variable given by $\lim_{n\to\infty} m_n$ and $\lim_{n\to\infty} \sigma_n$ respectively.

As a second application of theorem 1 we make the following observation. Denoting by $\{\xi(t)\}$ an arbitrary discrete stationary process, let a sequence (65) of processes be defined by

$$\{\xi^{(i)}(t)\} = [\ldots, \xi(t+1), \xi(t), \xi(t-1), \ldots, \xi(t-i), 0, 0, \ldots].$$

Then for every fixed $(t) = (t_1, .., t_n)$ we have $\lim_{i\to\infty} \xi^{(i)}(t_1, .., t_n) = = \xi(t_1, .., t_n)$. It follows that $\xi^{(i)}(t)$ converges in probability to $\{\xi(t)\}$.

Extending a current terminology, two processes, say $\{\xi^{(1)}(t)\}$ and $\{\xi^{(2)}(t)\}$, will be called *equivalent*, if for an arbitrary $(t) = (t_1, .., t_n)$ the two variables $\xi^{(1)}(t_1, .., t_n)$ and $\xi^{(2)}(t_1, .., t_n)$ are equivalent, i. e. if

$$P[\xi^{(1)}(t_1, .., t_n) \neq \xi^{(2)}(t_1, .., t_n)] = 0.$$

If two variables or processes are equivalent, we shall write $\xi^{(1)} = \xi^{(2)}$, $\{\xi^{(1)}(t)\} = \{\xi^{(2)}(t)\}$, etc.

14. On singular stationary processes.

Let $\xi = [\xi^{(1)}, .., \xi^{(n)}]$ represent an n-dimensional random variable with distribution function $F(u_1, .., u_n)$. The distribution of ξ will be called *linearly singular* or, more briefly, *singular*, if there exists a linear function, say $L[x-m] = a_1(x^{(1)} - m_1) + \cdots + a_n(x^{(n)} - m_n)$, such that

$$(68) \qquad P[L[\xi - m] \neq 0] = \\ = P[a_1(\xi^{(1)} - m_1) + \cdots + a_n(\xi^{(n)} - m_n) \neq 0] = 0.$$

If (68) is satisfied, the variables $\xi^{(i)}$ will be said to be connected by the relation $L[\xi - m] = 0$. The singularity will be said to be of *rank* h, if there exist $n-h$, and only $n-h$, independent relations between the variables $\xi^{(i)}$, say

$$(69) \qquad \begin{cases} a_{1,h+1} \cdot (\xi^{(1)} - m_1) + \cdots + a_{n,h+1} \cdot (\xi^{(n)} - m_n) = 0 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{1n} \cdot \quad (\xi^{(1)} - m_1) + \cdots + a_{nn} \quad \cdot \quad (\xi^{(n)} - m_n) = 0. \end{cases}$$

It will be relevant to interpret singularity in terms of characteristic functions. Denoting the characteristic function of ξ by $f(X_1, .., X_n)$, we have by definition

$$f(X_1, .., X_n) = E[e^{i(X_1 \cdot \xi^{(1)} + \cdots + X_n \cdot \xi^{(n)})}] =$$
$$= \int_{R_n} e^{i(X_1 \cdot x_1 + \cdots + X_n \cdot x_n)} \cdot d\,F(x_1, .., x_n).$$

Letting a matrix of real elements be given by

$$(70) \qquad \begin{pmatrix} a_{11}, & a_{12}, & .., & a_{1n} \\ a_{21}, & a_{22}, & .., & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1}, & a_{n2}, & .., & a_{nn} \end{pmatrix}$$

and writing $(Z) = (Z_1, .., Z_n)$ for an auxiliary set of real variables, let the substitution

$$(71) \qquad \begin{cases} X_1 = a_{11} \cdot Z_1 + \cdots + a_{1n} \cdot Z_n \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ X_n = a_{n1} \cdot Z_1 + \cdots + a_{nn} \cdot Z_n \end{cases}$$

transform $f(X_1, .., X_n)$ into $f^*(Z_1, .., Z_n)$. Considering the variable $\zeta = (\zeta^{(1)}, \ldots, \zeta^{(n)})$ defined by

$$(72) \qquad \begin{cases} \zeta^{(1)} = a_{11} \cdot \xi^{(1)} + \cdots + a_{n1} \cdot \xi^{(n)} \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \zeta^{(n)} = a_{1n} \cdot \xi^{(1)} + \cdots + a_{nn} \cdot \xi^{(n)}, \end{cases}$$

an elementary transformation shows that

$$(73) \quad f^*(Z_1, .., Z_n) = E[e^{i(X_1 \cdot \xi^{(1)} + \cdots + X_n \cdot \xi^{(n)})}] = E[e^{i(Z_1 \cdot \zeta^{(1)} + \cdots + Z_n \cdot \zeta^{(n)})}].$$

Thus f^* is nothing else than the characteristic function of the composite variable ζ. Now, let ξ be singular of rank h, say on account of the relations (69), and introduce for a moment the inconsequential assumption $m_i = 0$. In such case $f^*(Z_1, .., Z_n)$ will contain at most the variables $Z_1, .., Z_h$. On the other hand, if the matrix (70) is non-singular the first h variables must all appear in $f^*(Z_1, .., Z_n)$. In fact, since the distribution of ζ is uniquely determined by its

characteristic function (see e. g. H. CRAMÉR (1937) p. 104), the vanishing of Z_m in f^* for a number $m \leq h$ would imply $\zeta^{(m)} = 0$, a linear relation not obtainable from (69).

It will be observed that formula (73) holds for arbitrary values of the real coefficients a_{ik}. In particular, the characteristic function of the $h < n$ first variables $\zeta^{(i)}$ as given by (72) will be arrived at by putting $a_{ik} = 0$ for $k = h + 1,\ h + 2, \ldots, n$.

If ξ is singular on account of a relation of type (68), and if $E[\xi^{(i)}]$ exists, we shall always assume that

$$m_i = E[\xi^{(i)}], \tag{74}$$

which evidently will involve no loss of generality. Now, calculating $E[\zeta^{(i)}]$ from (72), and paying regard to (74), we get

$$\begin{cases} m_1^* = a_{11}\, m_1 + \cdots + a_{n1}\, m_n = E[\zeta^{(1)}] \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ m_n^* = a_{1n} m_1 + \cdots + a_{nn}\, m_n = E[\zeta^{(n)}]. \end{cases} \tag{75}$$

Next, let it be assumed that all the variables $\xi^{(i)}$ have a finite dispersion. Then the expectation

$$E\,[[X_1(\xi^{(1)} - m_1) + \cdots + X_n(\xi^{(n)} - m_n)]^2] =$$
$$= \int_{R_n} [X_1(x_1 - m_1) + \cdots + X_n(x_n - m_n)]^2 \cdot d\,F(x_1, \ldots, x_n)$$

will exist, and represent a non-negative quadratic form in the real variables X_i, say $Q(X_1, X_2, \ldots, X_n)$. Writing

$$\mu_{ik} = E\,[(\xi^{(i)} - m_i)(\xi^{(k)} - m_k)],$$

we have $\mu_{ik} = \mu_{ki}$, and

$$Q(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} \sum_{k=1}^{n} \mu_{ik} \cdot X_i X_k. \tag{76}$$

The quadratic form Q will give information concerning the singularity of the distribution of ξ. In fact, *the rank of the quadratic form equals the rank of the distribution.* In other words, by a suitable substitution in Q of type (71) the number of variables may be brought down to the rank of ξ, but not further — and *vice versa.* Denoting the transformed forms by Q^*, we have also

$$Q^*(Z_1, \ldots, Z_n) = E\,[[Z_1(\zeta^{(1)} - m_1^*) + \cdots + Z_n(\zeta^{(n)} - m_n^*)]^2],$$

so the forms Q^* thus are quite analogous to the characteristic functions f^* in respect of linear substitutions. As the proofs, too, may be given a parallel wording, the verification of the statements made needs no comment.[5]

After these preliminaries, the next step will be to investigate as to singularity the variables $\xi(t_1, \ldots, t_n)$ connected with a stationary process $\{\xi(t)\}$. Until further notice, we shall not assume that $\{\xi(t)\}$ has a finite dispersion.

Suppose first that $\xi(t, t-1, \ldots, t-n)$ is singular on account of the relation

$$\xi(t) - m + a_1 \cdot (\xi(t-1) - m) + \cdots + a_h \cdot (\xi(t-h) - m) = 0, \tag{77}$$

where $h \leq n$. Then (a) the relation of singularity (77) must hold for every t, (b) if $h < n$, the variable $\xi(t, t-1, \ldots, t-h)$ must also present the same singularity. Further, if $\xi(t, t-1, \ldots, t-h)$ satisfies a second linear relation of type (77), say with coefficients $(a_1', \ldots, a_h')$, we obtain by subtraction a linear relation showing that there is a $h' < h$ such that $\xi(t, t-1, \ldots, t-h')$ is singular. Thus, when considering a stationary process $\{\xi(t)\}$, a number h will be well-defined by terming $\{\xi(t)\}$ *singular of rank* h if in the sequence

$$\xi(t), \quad \xi(t, t-1), \quad \xi(t, t-1, t-2), \quad \ldots, \quad \xi(t, t-1, \ldots, t-n), \quad \ldots$$

the variable $\xi(t, t-1, \ldots, t-h)$ is the first singular one. Taking (77) for the singularity relation of $\xi(t, t-1, \ldots, t-h)$, it is readily seen that

(A) the coefficients a_i are uniquely determined, and $a_h \neq 0$.

(B) for all $k \geq 0$ the $(k+h+1)$-dimensional variable $\xi(t_0 + k, t_0 + k - 1, \ldots, t_0 - h)$ is singular of rank h, and a system of $k+1$ independent relations is given by

$$\begin{cases} \xi(t_0) - m + a_1 \cdot (\xi(t_0 - 1) - m) + \cdots + a_h \cdot \\ \qquad \cdot (\xi(t_0 - h) - m) = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \xi(t_0 + k - 1) - m + a_1 \cdot (\xi(t_0 + k - 2) - m) + \cdots + a_h \cdot \\ \qquad \cdot (\xi(t_0 + k - h - 1) - m) = 0 \\ \xi(t_0 + k) - m + a_1 \cdot (\xi(t_0 + k - 1) - m) + \cdots + a_h \cdot \\ \qquad \cdot (\xi(t_0 + k - h) - m) = 0 \end{cases} \tag{78}$$

Thus, if (77) holds the variable $\zeta(t_1, .., t_n)$ defined by (61), taking $a_0 = 1$, will reduce to $[E[\zeta], .., E[\zeta]]$. Further we note that the relations of singularity (78) extend so as to give the identity

$$(79) \qquad \xi(t) - m + a_1 \cdot (\xi(t-1) - m) + \cdots + a_h(\xi(t-h) - m) = 0$$

with $t = 0, \pm 1, \pm 2, \ldots$

According to observation (B), a sample value

$$\xi_i(t_0 + k, t_0 + k - 1, .., t_0 - h) = [\xi_i(t_0 + k), \xi_i(t_0 + k - 1), .., \xi_i(t_0 - k)]$$

will, when regarding $\xi_i(t_0 + t)$ as a function of t, with probability 1 satisfy the difference equation (32). Thus, if this difference equation reduces to (30), any sample series

$$[\xi_i(t_0 + k), \ \xi_i(t_0 + k - 1), .., \xi_i(t_0 - h)]$$

will with probability 1 be of type (17), i.e. consist of a number of superposed harmonics. This case will accordingly be referred to as a *process of superposed harmonics.*

Resuming the assumption that the process considered has a finite dispersion, we are now in a position to state

Theorem 2. Let $\{\xi(t)\}$ *be a discrete stationary process with autocorrelation coefficients* r_k. *If* $\{\xi(t)\}$ *is linearly singular,* $\{\xi(t)\}$ *is a process of superposed harmonics. A necessary and sufficient condition that* $\{\xi(t)\}$ *be linearly singular, say on account of the relation* $L[\xi(t) - m] = 0$ *given by* (77), *is that* r_k *satisfies the difference equation* $L[r_k] = 0$.

Denoting by m and σ the mean and dispersion of the process considered, the autocorrelation coefficients are given by (cf. (56))

$$(80) \qquad r_k = r_{-k} = E[(\xi(t) - m)\ (\xi(t-k) - m)]/\sigma^2.$$

Thus the quadratic form of type (76) belonging to $\xi(t, t-1, .., t-n)$, say $Q_n(X_t, X_{t-1}, .., X_{t-n})$, will be well-defined, and given by

$$Q_n(X_t, .., X_{t-n}) = \sigma^2 \cdot \sum_{p=0}^{n} \sum_{q=0}^{n} r_{|p-q|} \cdot X_{t-p} X_{t-q}.$$

Since the form Q_n is non-negative definite, its principal determinant will be non-negative (see e.g. G. Kowalewski (1909), Chapter 12),

$$(81) \qquad \Delta(r, n) = \begin{vmatrix} 1, & r_1, & r_2, & \ldots, & r_n \\ r_1, & 1, & r_1, & \ldots, & r_{n-1} \\ r_2, & r_1, & 1, & \ldots, & r_{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_n, & r_{n-1}, & r_{n-2}, & \ldots, & 1 \end{vmatrix} \geq 0.$$

The determinants $\Delta(r, n)$ defined above will be called the *principal correlation determinants* of the stationary process examined.

Thus prepared, let us begin with proving the second part of the theorem. In the first place, let $\{\xi(t)\}$ be singular on account of (77), and multiply the left member of this relation by $\xi(t-k)/\sigma^2$. Observing that the expectation of the resulting expression is zero, and paying regard to (80), we get

$$(82) \qquad L[r_k] = r_k + a_1 r_{k-1} + \cdots + a_h r_{k-h} = 0.$$

This relation shows that the condition is necessary.

On the other hand, let the autocorrelation coefficients r_t satisfy a linear difference equation $L[r_t] = 0$ of order h^*. Transforming this equation to the form (32), and reducing to lowest possible order, say h, let $L_1[r_t] = 0$ be the result. According to the previous analysis, $\xi(t)$ will then satisfy no linear difference relation of order $< h$. After this remark, let the consecutive rows of $\Delta(r, h)$ be denoted by $\varrho_0, \varrho_1, \ldots, \varrho_h$. From the structure of $\Delta(r, h)$ it is evident that these rows are connected by the linear relation $L_1[\varrho_h] = 0$. Thus $\Delta(r, h)$ equals zero, so the rank of Δ will be $\leq h$. Recalling from the theory of quadratic forms that the rank of $\Delta(r, n)$ equals the rank of Q_n, and keeping in mind the identity between the ranks of corresponding distributions and quadratic forms of type (76), it turns out that $\xi(t, t-1, \ldots, t-h)$ is linearly singular. Let the relation of singularity be $L_2[\xi(t) - m] = 0$. Now, were not $L_2 \equiv L_1$, then, contrarily to the assumption made, r_t would satisfy a linear difference relation of order $< h$. Moreover, according to the construction of L_1 we have $L[r_k] = L^*[L_1[r_k]]$, where L^* is a well-defined linear operation. Thus $L[\xi(t) - m] = L^*[L_1[\xi(t) - m]] = = L^*[0] = 0$, which proves that the condition is sufficient.

In order to prove the first part of theorem 2, let (77) be the relation of singularity, and let it be assumed that this has already been reduced to the lowest possible order. According to the above, r_k will then satisfy (82), but no linear equation of lower order.

Writing r_t on the form (33), this implies that none of the polynomials H will be vanishing identically. Since, finally, the inequality $|r_k| \leq 1$ shows that r_k is uniformly bounded in modulus in $(-\infty < k < \infty)$, we conclude from the second remark in section 6 that r_k can be written on the form (17).

After this analysis, the following theorem will need no comment.

Theorem 3. Let $\{\xi(t)\}$ be a discrete stationary process with principal correlation determinants $\Delta(r, n)$ given by (81). *A necessary and sufficient condition that $\{\xi(t)\}$ be singular of rank h is that $\Delta(r, h)$ be the first vanishing determinant in the sequence $\Delta(r, 1), \Delta(r, 2), \ldots$*

A relation of type (77) will be called a *stochastical difference relation of order h* satisfied by the stationary process $\{\xi(t)\}$. The previous analysis shows that a stationary process which has finite dispersion and satisfies (77) will satisfy a difference relation of type

$$(83) \quad \Delta^{2s}\xi(t-s) + h_1 \Delta^{2s-2}\xi(t-s+1) + \cdots + h_s[\xi(t) - m] = 0.$$

In the next two sections it will be shown, i. a., that the theorems of the present section are not vacuous, i. e. that there really exist stationary processes having the properties assumed by hypothesis.

15. Some type cases of the discrete stationary process.

As mentioned in section 12, it will be shown in the present section that the schemes surveyed in Chapter I may be regarded as special cases of the discrete stationary process. Some other schemes will also be presented, and a few characteristic properties of the different types considered be pointed out. Conditioned variables and expectations, and the operation of addition will be exemplified, and devices given for the construction of model series which follow the different schemes. For further concreteness some model series constructed for the illustration of later results will be furnished.

α. The purely random process. This term will in the sequel be used for the purely random scheme touched upon in section 7. The purely random process will, of course, be obtained by taking, in the relation (52) defining the general process,

$$F(t_1, \ldots, t_n;\ u_1, \ldots, u_n) = F(u_1) \ldots F(u_n),$$

where any distribution function may be chosen for $F(u)$. The verification of (53)—(55) is obvious.

When detailed information is required, a purely random process $\{\xi(t)\}$ defined by a distribution function $F(u)$ will be denoted by $\{\xi(t;\ F)\}$. It is seen that the defining function $F(u)$ is identical with the principal distribution function of the process.

The following simple theorem exemplifies the operation of addition of independent processes.

Theorem 4. Let $\{\xi^{(1)}(t;\ F^{(1)})\}$, $\{\xi^{(2)}(t;\ F^{(2)})\}$, ... *represent independent, purely random processes such that the infinite convolution*

$$F^{(1)} * F^{(2)} * \ldots. \tag{84}$$

is convergent. Then the sum $\{\xi^{(1)}(t;\ F^{(1)})\} + \{\xi^{(2)}(t;\ F^{(2)})\} + \cdots$ *will be convergent, and constitute a purely random process with the convolution* (84) *for principal distribution function.*

In fact, the convergent convolution (84) is the distribution function of the sum $\sum_{i=1}^{\infty} \xi^{(i)}(t;\ F^{(i)})$, which is thus convergent. According to a remark attached to theorem 1, the convergence of this sum implies the convergence of $\sum_{i=1}^{\infty} \{\xi^{(i)}(t;\ F^{(i)})\}$.

A characteristic property of the purely random process is that the two variables $\xi_C(t_1, \ldots, t_n)$ and $\xi(t_1, \ldots, t_n)$ will have identical distribution functions for any condition (C) not referring to any time point in the set $(t) = (t_1, \ldots, t_n)$. Thus we have in this case (cf. (57))

$$E_C[\xi(t_{m+1})] = E[\xi(t)].$$

Any random series, e. g. a series of records on throws with a die, will form a model series of the purely random process. The illustrative model series used in the present study are very simply constructed, the double purpose being to facilitate the calculations, and to bring into relief the characteristic features of the different types of process. For this construction, the well-known random sampling numbers of L. H. C. Tippett (1927) were used.

Two independent random series, denoted by $(\bar{a}_t^{(1)})$ and $(\bar{a}_t^{(2)})$, will be given for illustration. Denoting the corresponding processes by $\{\alpha^{(1)}(t;\ F_1)\}$ and $\{\alpha^{(2)}(t;\ F_2)\}$ respectively, the defining function F_1 is given by

$$F_1(u) = 0 \text{ for } u < -1, \qquad F_1(u) = \cdot 1 \text{ for } -1 \leq u < 0,$$
$$F_1(u) = \cdot 9 \text{ for } 0 \leq u < 1, \quad F_1(u) = 1 \text{ for } 1 \leq u,$$

and the function F_2 by

$$F_2(u) = 0 \text{ for } u < -1, \qquad F_2(u) = \cdot 3 \text{ for } -1 \leq u < 0,$$
$$F_2(u) = \cdot 7 \text{ for } 0 \leq u < 1, \quad F_2(u) = 1 \text{ for } 1 \leq u.$$

A short calculation shows that the mean value of each process $\{\alpha^{(i)}\}$ equals zero, and that the variances, say D_1^2 and D_2^2, are

(85) $$D_1^2 = \cdot 2, \qquad D_2^2 = \cdot 6.$$

Writing τ for the TIPPETT numbers, the model series $(\bar{\alpha}_t^{(1)})$ consists of the 1000 elements obtained from the first 1000 τ-numbers on page 1 by the use of the following code:

$$\bar{\alpha}_t^{(1)} = 1 \text{ for } \tau = 0; \quad \bar{\alpha}_t^{(1)} = 0 \text{ for } \tau = 1, \ldots, 8; \quad \bar{\alpha}_t^{(1)} = -1 \text{ for } \tau = 9.$$

The second series was obtained from the corresponding τ-numbers on page 2. The code used was

$$\bar{\alpha}_t^{(2)} = 1 \text{ for } \tau = 0, 1, 2; \quad \bar{\alpha}_t^{(2)} = 0 \text{ for } \tau = 3, \ldots, 6;$$
$$\bar{\alpha}_t^{(2)} = -1 \text{ for } \tau = 7, 8, 9.$$

The first 100 elements in each of the $\bar{\alpha}$-series will now be quoted.

Table 1. (1) *Model series* $(\bar{\alpha}_t^{(1)})$; *first 100 elements.*

0	−1	0	0	0	0	0	0	0	−1	−1	0	−1	0	−1	0	0	−1	0	−1
0	−1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	−1	0	0	0
0	0	0	0	0	0	−1	0	0	0	1	0	0	0	0	0	0	0	1	1
0	0	−1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	−1	0	−1	0	0	0	0	−1	0	1	0	0	1

(2) *Model series* $(\bar{\alpha}_t^{(2)})$; *first 100 elements.*

1	1	0	0	1	−1	0	−1	−1	0	0	−1	0	1	1	0	0	0	−1	0
−1	0	−1	0	1	0	1	−1	0	0	−1	−1	0	0	0	0	0	−1	1	1
1	−1	1	1	−1	1	0	0	0	−1	−1	−1	1	0	1	1	−1	−1	1	1
1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	1	−1	−1	−1	−1
0	−1	−1	0	0	0	1	1	1	−1	0	0	1	0	−1	0	−1	−1	1	−1

The $\bar{\alpha}$-series will be used repeatedly in the present study. For this reason each of the series, considered a statistical population with hypothetical distribution functions $F_1(u)$ and $F_2(u)$ respectively, has been tested as to the goodness of fit between the empirical and the theoretical distributions. Denoting the empirical distribution functions by $\overline{F}_1(u)$ and $\overline{F}_2(u)$ respectively, these were found to be

$$(86)\qquad \begin{cases} \overline{F}_1(u) = 0 \text{ for } u < -1, & \overline{F}_1(u) = \cdot 100 \text{ for } -1 \leq u < 0, \\ \overline{F}_1(u) = \cdot 908 \text{ for } 0 \leq u < 1, & \overline{F}_1(u) = 1 \text{ for } 1 \leq u; \end{cases}$$

$$(87)\qquad \begin{cases} \overline{F}_2(u) = 0 \text{ for u} < -1, & \overline{F}_2(u) = \cdot 321 \text{ for } -1 \leq u < 0, \\ \overline{F}_2(u) = \cdot 692 \text{ for } 0 < u < 1, & \overline{F}_2(u) = 1 \text{ for } 1 \leq u. \end{cases}$$

The ω^2-test (see p. 21) indicates a nice fit. As is readily verified by the insertion of (86) and (87), the two $\bar{\alpha}$-series give the ω^2-values ˙000064 and ˙000505, while the corresponding expectations are ˙000180 and ˙000420.

The model series $(\bar{\alpha})$ have also been tested with regard to the concordance between serial $(\bar{r}_k)$ and autocorrelation (r_k) coefficients. The latter vanish for $k \neq 0$. On the other hand, writing n for the number of elements in the series, and paying no regard to terms of order $1/n$, the sampling dispersion of any autocorrelation coefficient of the two series is found to be $1/\sqrt{n} = \cdot 032$. The first five serial coefficients are given below.

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$\bar{r}_k(\bar{\alpha}_t^{(1)})$	˙057	˙047	˙010	˙015	˙036
$\bar{r}_k(\bar{\alpha}_t^{(2)})$	˙046	˙011	˙006	−˙004	−˙004

The deviation from the corresponding autocorrelation coefficient is in no case larger than the double dispersion. It is rather interesting to note that although the series consist of as many as 1000 elements, a serial coefficient amounting to ˙06 cannot be considered significantly positive.

The first 20 values of the series $(\bar{\alpha}_t^{(1)} - \bar{\alpha}_t^{(2)})$ obtained from table 1 are given below to illustrate a linear operation with independent random processes:

(88) −1 −2 0 0 −1 1 0 1 1 −1 −1 1 −1 −1 −2 0 0 −1 1 −1

The series thus obtained is a model series of the process $\{\alpha_t^{(1)} - \alpha_t^{(2)}\}$. Since in the present case $\{\alpha_t^{(1)} + \alpha_t^{(2)}\}$ is found to possess the same defining distribution function as $\{\alpha_t^{(1)} - \alpha_t^{(2)}\}$, the series (88) also forms a model series for the process $\{\alpha_t^{(1)} + \alpha_t^{(2)}\}$.

We shall next pass to some other type cases of the discrete stationary process, denoted by β, γ and δ, which will be built up by successive linear operations on the purely random processes.

β. *The process of moving averages.* According to the analysis in section 13, a stationary process $\{\xi(t)\}$ will be obtained by taking

$$(89) \qquad \xi(t) = b_0 \cdot \eta(t) + b_1 \cdot \eta(t-1) + \cdots + b_h \cdot \eta(t-h),$$

letting $\{\eta(t)\}$ represent a purely random process, say $\{\eta(t; F_\eta(u))\}$, and $(b) = (b_0, b_1, \ldots, b_h)$ an arbitrary sequence of real numbers. The type of process thus defined will be called *the process of moving averages.* A specific process (89) of this type will sometimes be denoted by $\{\xi(t; \eta)\}$. The purely random process $\{\eta(t)\}$ and the variables $\eta(t_1, \ldots, t_n)$ will be called *primary* in respect of $\{\xi(t; \eta)\}$ and $\xi(t_1, \ldots, t_n)$ respectively.

Let it be observed that, for any constant $c > 0$, the variable (89) is identical with that defined by the variables $\eta(t; F_\eta(c \cdot u))$ and by the sequence $c \cdot (b) = (c \cdot b_0, c \cdot b_1, \ldots, c \cdot b_h)$. Therefore, the assumption $b_0 = 1$ often imposed on (89) in the following will not restrict the generality of the analysis. On the other hand, if $\Sigma b_i \neq 0$, we can find an identical process such that $\Sigma b_i = 1$. Hence the name proposed for the process.

The principal distribution functions $F_\xi(u)$ and $F_\eta(u)$ of $\{\xi(t)\}$ and $\{\eta(t)\}$ respectively are connected by the relation

$$(90) \qquad F_\xi(u) = F_\eta(u/b_0) * F_\eta(u/b_1) * \cdots * F_\eta(u/b_h).$$

In case $D(\eta(t))$ is finite, we obtain further

$$D^2(\xi(t)) = (b_0^2 + b_1^2 + \cdots + b_h^2) \cdot D^2(\eta(t)).$$

For exemplifying conditioned variables connected with a process $\{\xi(t; \eta)\}$ of moving averages, let $m < h$, and let C denote the condition

$$(C) = (\eta(t-m-1) = \eta_{m+1},\ \eta(t-m-2) = \eta_{m+2}, \ldots, \eta(t-h) = \eta_h).$$

Then the conditioned variable $\xi_C(t)$ will be given by

$$\xi_C(t) = b_0 \cdot \eta(t) + \cdots + b_m \cdot \eta(t-m) + b_{m+1} \cdot \eta_{m+1} + \cdots + b_h \cdot \eta_h \cdot$$

As is readily verified, we have in case $D(\eta(t))$ is finite,

$$(91) \quad E_C[\xi(t)] = (b_0 + b_1 + \cdots + b_m) E[\eta(t)] + b_{m+1} \cdot \eta_{m+1} + \cdots + b_h \cdot \eta_h \cdot$$

Model series for the process of moving averages are readily obtained by applying a moving linear operation to a model series for the purely random process. For instance, the series of differences $\Delta^k \bar{\eta}_t$, of any order k, obtained from a purely random series $\bar{\eta}_t$ will illustrate the process of moving averages. Below, as an illustrative model series are given the first 100 values of $\Delta \bar{\alpha}_t^{(2)} = \bar{\alpha}_{t+1}^{(2)} - \bar{\alpha}_t^{(2)}$ obtained from Table 1. The corresponding process will be denoted

(92) $$\{\beta(t; \alpha^{(2)})\} = \{\alpha^{(2)}(t+1; F_2) - \alpha^{(2)}(t; F_2)\}.$$

Table 2. Model series ($\bar{\beta}_t$); *first* 100 *elements.*

0	−1	0	1	−2	1	−1	0	1	0	−1	1	1	0	−1	0	0	−1	1	−1
1	−1	1	1	−1	1	−2	1	0	−1	0	1	0	0	0	0	−1	2	0	0
−2	2	0	−2	2	−1	0	0	−1	0	0	2	−1	1	0	−2	0	2	0	0
0	0	0	−1	1	−1	1	0	−1	1	0	−1	1	0	0	−2	0	0	0	1
−1	0	1	0	0	1	0	0	−2	1	0	1	−1	−1	1	−1	0	2	−2	2

γ. *The general process of linear regression.* Let $\{\eta(t)\}$ stand for a purely random process with finite dispersion $D(\eta)$, and let b_0, b_1, b_2, ... be a real sequence such that $\sum_{k=0}^{\infty} b_k^2$ is convergent. Finally, we must assume either that $E[\eta] = 0$ or that $\sum_{k=0}^{\infty} b_k$ be convergent. These alternatives will present themselves repeatedly in the sequel. The former assumption is better suited to our purpose; the modifications caused by the latter are trivial. Accordingly, we shall always assume that $E[\eta] = 0$.

Considering the series

(93) $$b_0\,\eta(t) + b_1\,\eta(t-1) + b_2\,\eta(t-2) + \ldots,$$

it follows from the independence of the variables $\eta(t)$ that the variance of

(94) $$b_n\,\eta(t-n) + b_{n+1}\,\eta(t-n-1) + \cdots + b_{n+p}\,\eta(t-n-p)$$

is given by

$$(b_n^2 + b_{n+1}^2 + \cdots + b_{n+p}^2) \cdot D^2(\eta).$$

The dispersion of (94) thus tends to zero uniformly in p as $n \to \infty$. Accordingly, the sum (93) is convergent (cf. p. 41). As remarked in connexion with theorem 1, it follows that we may define a stationary process $\{\xi(t)\}$ by writing

$$\{\xi(t)\} = b_0\{\eta(t)\} + b_1\{\eta(t-1)\} + b_2\{\eta(t-2)\} + \cdots \tag{95}$$

By definition, this is the general formula for a *process of linear regression*.

Having dealt under article β with a particular case of the general process of linear regression, we shall in the next article build up the scheme of autoregression as another type case. As a matter of fact, in the present study the general process will be chiefly used as a convenient tool for comprehension of the two type cases mentioned, and it is only these which will appear in the applications. It will also be sufficient to refer to the type cases for model series.

δ. *The process of linear autoregression.* Let $(a) = (a_1, \ldots, a_h)$ stand for real numbers such that $a_h \neq 0$, and that the roots of the equation (34) all are of a modulus less than 1. Let further $(b) = (b_1, b_2, \ldots)$ be a sequence such that (A) the difference equation

$$x(t) + a_1 \cdot x(t-1) + \cdots + a_h \cdot x(t-h) = 0 \tag{96}$$

is satisfied when $x(t) \equiv b_t$, and (B) the initial values $b_1, \ldots, b_h$ are solutions of the following system of linear equations

$$\left\{\begin{array}{l} a_1 \quad + b_1 = 0 \\ a_2 \quad + a_1 \cdot b_1 + b_2 = 0 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{h-1} + a_{h-2} \cdot b_1 + \cdots + a_1 \cdot b_{h-2} + b_{h-1} = 0. \\ a_h \quad + a_{h-1} \cdot b_1 + \cdots\cdots\cdots + a_1 \cdot b_{h-1} + b_h = 0. \end{array}\right. \tag{97}$$

Since the determinant of the system (97) is given by

$$\begin{vmatrix} 1 & 0 & 0 \cdots 0 & 0 \\ a_1 & 1 & 0 \cdots 0 & 0 \\ a_2 & a_1 & 1 \cdots 0 & 0 \\ \cdot & \cdot & \cdots\cdot & \cdot \\ \cdot & \cdot & \cdots\cdot & \cdot \\ a_{h-1} & a_{h-2} & \cdot \cdots a_1 & 1 \end{vmatrix}$$

and thus equals 1, the initial b-values are uniquely determined. It is also seen that all b_i are real. Moreover, the series $\sum_0^\infty b_s^2$ is convergent (cf. section 6). Letting $\{\eta(t)\}$ represent a purely random process with finite dispersion, the conditions indicated under article γ thus will be satisfied. Hence, a stationary process $\{\xi(t)\} = \{\xi(t;\eta)\}$ will be defined by putting

$$\xi(t;\eta) = \eta(t) + b_1 \cdot \eta(t-1) + b_2 \cdot \eta(t-2) + \cdots \tag{98}$$

By definition, this operation gives rise to the general *process of (linear) autoregression*. Since $a_h \neq 0$, the autoregression will be said to be of *order* h.

As pointed out in section 13, the variables $\zeta(t)$ defined by the following linear operation on the variables $\xi(t;\eta)$ given by (98) will likewise constitute a stationary process:

$$\zeta(t) = \xi(t) + a_1 \cdot \xi(t-1) + \cdots + a_h \cdot \xi(t-h).$$

It will now be shown that the process $\{\zeta(t)\}$ thus defined is equivalent with $\{\eta(t)\}$. The proof is based on a transformation of a double sum of aleatory variables.

By definition, we have

$$\zeta(t) = \sum_{i=0}^{h} a_i \sum_{k=0}^{\infty} b_k \cdot \eta(t-i-k) = \lim_{N\to\infty} \sum_{i=0}^{h} a_i \sum_{k=0}^{N} b_k \cdot \eta(t-i-k),$$

where a_0 and b_0 should be given the value 1. Introducing an auxiliary variable $\zeta_N(t)$, an elementary transformation shows that

$$\zeta_N(t) = \sum_{i=0}^{h} a_i \sum_{k=0}^{N} b_k \, \eta(t-i-k) = \sum_{p=0}^{h-1} \eta(t-p) \sum_{q=0}^{p} a_{p-q} \, b_q + \sum_{p=h}^{N} \eta(t-p) \sum_{q=0}^{h} a_q \, b_{p-q} +$$

$$+ \sum_{p=N+1}^{N+h} \eta(t-p) \sum_{q=p-N}^{h} a_q \, b_{p-q} = \eta(t) + c_1 \cdot \eta(t-N-1) + \cdots + c_h \cdot \eta(t-N-h),$$

where the second transformation is a consequence of (96) and (97). Putting $a = \max |a_i|$, the coefficients c_s introduced are seen to satisfy the inequality

$$|c_s| \leq a \cdot (|b_{N+s-h}| + |b_{N+s+1-h}| + \cdots + |b_{N-1}| + |b_N|). \tag{99}$$

Paying regard to the convergence of $\sum b_i^2$, we conclude without difficulty that $c_1 \, \eta(t-N-1) + \cdots + c_h \, \eta(t-N-h)$ tends to zero

in probability as $N \to \infty$. Thus, $\zeta_N(t)$ tends to $\eta(t)$ in probability. According to the corollary of theorem 1, the proof that $\{\zeta(t)\}$ equals $\{\eta(t)\}$ is thereby completed, and we get the following fundamental identity,

$$(100) \qquad \{\xi(t)\} + a_1 \cdot \{\xi(t-1)\} + \cdots + a_h \cdot \{\xi(t-h)\} = \{\eta(t)\}.$$

The relation (100) says that we have, for every $t = 0, \pm 1, \pm 2, \ldots$,

$$(101) \qquad \xi(t) + a_1 \cdot \xi(t-1) + \cdots + a_h \cdot \xi(t-h) = \eta(t).$$

Since by assumption $\eta(t)$ is independent of $\eta(t-1)$, $\eta(t-2), \ldots$, and — according to (98) — also of $\xi(t-1)$, $\xi(t-2), \ldots$, the relation (101) shows that the variables $\xi(t)$, $\xi(t-1), \ldots, \xi(t-h)$ are connected by a relation of linear regression. Hence the name proposed for the process.

A simple illustration of conditioned variables is given by

$$\xi_C(t) = \eta(t) - a_1 \xi_{t-1} - a_2 \xi_{t-2} - \cdots - a_h \xi_{t-h},$$

where $(C) = (\xi(t-1) = \xi_{t-1}, \quad \xi(t-2) = \xi_{t-2}, \ldots, \quad \xi(t-h) = \xi_{t-h})$. More general formulae will be given in section 23.

Construction of model series for a process of autoregression, given for example by (98), may be performed in the same way as in the case of a finite moving average of a purely random series. The difficulty of an infinite number of weights b_s is but apparent, for when a certain precision in the calculations is fixed, only a finite number of the weights, say H of them, will be found to have any influence.

Denoting by $\bar{\delta}_t$ the values in a model series of the present type, and representing the primary model series by $(\bar{a}_t)$, the formula for the construction reads (cf. (63))

$$\bar{\delta}_t = \bar{a}_t + b_1 \cdot \bar{a}_{t-1} + b_2 \cdot \bar{a}_{t-2} + \cdots + b_H \cdot \bar{a}_{t-H}.$$

Having constructed $\bar{\delta}_1, \bar{\delta}_2, \ldots, \bar{\delta}_H$ according to this formula, the subsequent values $\bar{\delta}_{H+1}$, $\bar{\delta}_{H+2}$, etc. may be obtained from the more convenient recurrence formula

$$(102) \qquad \bar{\delta}_t = \bar{a}_t - a_1 \cdot \bar{\delta}_{t-1} - a_2 \cdot \bar{\delta}_{t-2} - \cdots - a_h \cdot \bar{\delta}_{t-h}.$$

In the illustrative series given below, a slight simplification has been made, in that formula (102) has been applied also for $t = 1$,

2, . ., H, taking $\bar{\delta}_t = 0$ for $t < 1$. As the series consist of 1000 elements each, this modification of the first few elements will not have any disturbing effect upon serial coefficients and other quantities relating to the whole of the series. On the other hand, thanks to the modification adopted, the construction of the model series may be readily followed in detail.

Three illustrative series, denoted by $(\bar{\delta}_t^{(1)})$, $(\bar{\delta}_t^{(2)})$ and $(\bar{\delta}_t^{(3)})$ respectively, will be presented. The formulae of type (100) for the corresponding processes read

(103) $$\{\delta^{(1)}(t)\} = \{\alpha^{(2)}(t)\} - \cdot 8\,\{\delta^{(1)}(t-1)\},$$

(104) $$\{\delta^{(2)}(t)\} = \{\alpha^{(1)}(t)\} + \cdot 8\,\{\delta^{(2)}(t-1)\},$$

(105) $$\{\delta^{(3)}(t)\} = \{\alpha^{(2)}(t)\} + \cdot 2\,\{\delta^{(3)}(t-1)\} - \cdot 65\,\{\delta^{(3)}(t-2)\}.$$

The verification of the recurrent calculation of the $\bar{\delta}$-series needs no comment.

Table 3. (1) *Model series* $(\bar{\delta}_t^{(1)})$. *First 50 elements.*

1·00	·20	− ·16	·13	·90	−1·72	1·37	−2·10	·68	− ·54
·44	−1·35	1·08	·14	·90	− ·72	·57	− ·46	− ·64	·51
−1·41	1·12	−1·90	1·52	− ·22	·17	·86	−1·69	1·35	−1·08
− ·13	− ·89	·71	− ·57	·46	− ·37	·29	−1·23	1·99	− ·59
1·47	−2·18	2·74	−1·19	− ·04	1·04	− ·83	·66	− ·53	− ·58

(2) *Model series* $(\bar{\delta}_t^{(2)})$. *First 50 elements.*

·00	−1·00	− ·80	− ·64	− ·51	− ·41	− ·33	− ·26	− ·21	−1·17
−1·93	−1·55	−2·24	−1·79	−2·43	−1·95	−1·56	−2·25	−1·80	−2·44
−1·95	−2·56	−2·05	−1·64	−1·31	−1·05	− ·84	·33	·26	·21
·17	·14	·11	·09	·07	·06	− ·95	− ·76	− ·61	− ·49
− ·39	− ·31	− ·25	− ·20	− ·16	− ·13	−1·10	− ·88	− ·71	− ·56

(3) *Model series* $(\bar{\delta}_t^{(3)})$. *First 50 elements.*

1·00	1·20	− ·41	− ·86	1·09	− ·22	− ·76	−1·01	− ·71	·51
·56	−1·22	− ·61	1·67	1·73	− ·74	−1·27	·23	− ·13	− ·17
− ·95	− ·08	− ·40	− ·03	1·25	·27	·24	−1·13	− ·38	·66
− ·62	−1·55	·09	1·03	·15	− ·64	− ·22	− ·63	1·02	1·61
·66	−1·92	·19	2·28	− ·67	− ·62	·31	·47	− ·11	−1·32

π. *On the periodic processes.* In this article, a stationary process will be constructed which belongs to the class of singular processes as introduced in section 14. A distribution function $F(u)$ and an integer h being arbitrarily given, there will be constructed a stationary process, say $\{\xi(t)\}$, with the following properties: (a) the process will be singular by means of the relation

$$\xi(t) - \xi(t-h) = 0, \tag{106}$$

(b) the process has $F(u)$ for principal distribution function. According to the construction, any sample series, say $(\xi_t, \xi_{t+1}, \ldots)$, will be strictly periodic, and with period h. The process will, accordingly, be termed a *periodic process.*

The simple construction device reads as follows. Let $\xi = [\xi^{(1)}, \ldots, \xi^{(h)}]$ represent an h-dimensional aleatory variable such that (A) the distribution function of ξ, say $F(u_1, \ldots, u_n)$, is symmetrical in respect of the variables u_i, (B) all the variables $\xi^{(i)}$ have $F(u)$ for distribution function. Now, let a sequence of multi-dimensional aleatory variables be defined by

$$\begin{array}{lll}
[\xi^{(1)}], & [\xi^{(1)}, \xi^{(2)}], \ldots\ldots\ldots\ldots, & [\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(h)}], \\
[\xi^{(1)}, \ldots, \xi^{(h)}, \xi^{(1)}], & [\xi^{(1)}, \ldots, \xi^{(h)}, \xi^{(1)}, \xi^{(2)}], \ldots, & [\xi^{(1)}, \ldots, \xi^{(h)}, \xi^{(1)}, \ldots, \xi^{(h)}], \\
\multicolumn{3}{l}{[\xi^{(1)}, \ldots, \xi^{(h)}, \xi^{(1)}, \ldots, \xi^{(h)}, \xi^{(1)}], \ldots\ldots\ldots\ldots\ldots\ldots} \\
\multicolumn{3}{l}{\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots}
\end{array}$$

It is evident that these variables may be taken for the variables

$$\begin{array}{l}
\xi(t), \qquad \xi(t, t-1), \ldots\ldots, \xi(t, t-1, \ldots, t-h+1), \\
\xi(t, \ldots, t-h+1, t-h), \xi(t, \ldots, t-h+1, t-h, t-h-1), \ldots, \\
\qquad\qquad \xi(t, \ldots, t-h+1, t-h, \ldots, t-2h+1), \\
\xi(t, \ldots, t-h+1, t-h, \ldots, t-2h+1, t-2h), \ldots\ldots\ldots \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{array}$$

connected with a stationary process $\{\xi(t)\}$. It needs no comment that this process has the advanced properties (a) and (b).

For exemplifying conditioned probability distributions connected with the periodic process constructed above, let C stand for a condition implying $\xi(t) = \xi_t$, and let n represent an integer. Then $P_C[\xi(t + n \cdot h) \subset \theta] = 1$ if ξ_t belongs to the set θ, otherwise $P_C = 0$. Further, $E_C[\xi(t + n \cdot h)] = \xi_t$.

For the construction of a model series, say $\bar{\pi}_1, \bar{\pi}_2, \ldots$, illustrating the periodic process defined above, it will be sufficient to form a

model random sample $\bar{\pi}_1, \bar{\pi}_2, \ldots, \bar{\pi}_h$ on the basis of the distribution function $F(u)$, and then take $\bar{\pi}_{h+1} = \bar{\pi}_1$, $\bar{\pi}_{h+2} = \bar{\pi}_2$, etc. Still simpler, a model series with $h = 2$ has been obtained by letting a coin-throw decide whether $\bar{\pi}_1$ should equal 1 or -1, and by taking $\bar{\pi}_{2i} = -\bar{\pi}_1$, $\bar{\pi}_{2i+1} = \bar{\pi}_1$. The resulting series reads

$$(107) \quad 1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1 \quad 1 \quad \ldots\ldots$$

This series will be referred to as the $\bar{\pi}$-series.

Considering the difference equation (14) corresponding to the relation of singularity (106) of the periodic process constructed, this equation has for general solution a sum of harmonics, viz. the expression (15), a circumstance in full agreement with theorem 2. This aspect corresponds to interpreting a sample series π_t by means of FOURIER analysis as a composed harmonic (cf. section 5).

Having so far shown that the class of singular stationary processes introduced in section 14 is not vacuous, the next question is if there exist singular processes which satisfy relations (77) of a more general type than (106). As a matter of fact, any particular relation generalizing (106) will impose certain conditions upon the distribution functions $\{F\}$ of the singular process. Leaving open the question of which special distribution functions may present themselves in case of a special singularity of type (77), we advance that the normal process (see section 16) will be found to admit any relation (83) as a singular case. After this reference to a process of superposed harmonics, only one of the schemes surveyed in Chapter I, viz. the scheme of hidden periodicities, remains to be interpreted as a stationary process.

Ω. The process of hidden periodicities. Let $\{\xi^{(1)}(t)\}, \ldots, \{\xi^{(k)}(t)\}$ represent independent stationary processes. According to section 13, the sum $\{\xi(t)\} = \{\xi^{(1)}(t)\} + \cdots + \{\xi^{(k)}(t)\}$ will constitute a stationary process. If at least one of the processes $\{\xi^{(i)}(t)\}$ is a periodic process, or a process of superposed harmonics, $\{\xi(t)\}$ will be called a *process of hidden periodicities.* In particular, letting $k = 2$, and taking for $\{\xi^{(1)}(t)\}$ a process of superposed harmonics, and for $\{\xi^{(2)}(t)\}$ a purely random process, we get the simple scheme of hidden periodicities dealt with in section 8.

A model series for the process of hidden periodicities may be obtained from independently constructed model series for the purely random process and the periodic process. Taking one model series

of each type, the series obtained by summing corresponding elements will form a model series for the process of hidden periodicities. The table below gives the first elements in two model series for the processes $\{\Omega^{(1)}\}$ and $\{\Omega^{(2)}\}$ defined by

$$\Omega^{(i)}(t) = \pi(t) + \alpha^{(i)}(t),$$

the α- and π-processes being defined in corresponding articles of the present section. The two $\bar{\Omega}$-series consist of 1000 elements each. The construction may be followed in detail by means of table 1 and the series (107) (cf. formula (39)).

Table 4. (1) *Model series* $(\bar{\Omega}_t^{(1)})$; *first* 100 *elements.*

1	−2	1	−1	1	−1	1	−1	1	−2	0	−1	0	−1	0	−1	1	−2	1	−2
1	−2	1	−1	1	−1	1	0	1	−1	1	−1	1	−1	1	−1	0	−1	1	−1
1	−1	1	−1	1	−1	0	−1	1	−1	2	−1	1	−1	1	−1	1	−1	2	0
1	−1	0	−1	1	−1	1	0	1	−1	1	−1	1	−1	2	−1	1	−1	1	−1
1	−1	1	−1	1	0	1	−2	1	−2	1	−1	1	−1	0	−1	2	−1	1	0

(2) *Model series* $(\bar{\Omega}_t^{(2)})$; *first* 100 *elements.*

2	0	1	−1	2	−2	1	−2	0	−1	1	−2	1	0	2	−1	1	−1	0	−1
0	−1	0	−1	2	−1	2	−2	1	−1	0	−2	1	−1	1	−1	1	−2	2	0
2	−2	2	0	0	0	1	−1	1	−2	0	−2	2	−1	2	0	0	−2	2	0
2	0	2	0	1	0	1	0	2	−1	2	0	1	0	2	0	0	−2	0	−2
1	−2	0	−1	1	−1	2	0	2	−2	1	−1	2	−1	0	−1	0	−2	2	−2

To give an example of conditional variables and expectations in the case of hidden periodicities, we note the simple relations

$$\Omega_C(t + n \cdot h) = \pi_t + \xi(t + n \cdot h); \qquad E_C[\Omega(t + n\,h)] = \pi_t + E[\xi].$$

Here $\{\pi(t)\}$ and $\{\xi(t)\}$ represent independent stationary processes with sum $\{\Omega(t)\}$. The process $\{\pi(t)\}$ is assumed to be periodic with period h, while n is an arbitrary integer, and C stands for the condition $(C) = (\pi(t) = \pi_t)$.

The scheme of hidden periodicities (39) and its many important applications are well-known from the text-books of time-series analysis. We have now interpreted this scheme as a special case of the stationary process. In what follows the scheme (39) with its rigid periodicities will be touched upon only incidentally. Our main theme is the equally important but largely unexplored processes of linear regression.

16. On the normal stationary process.

With reference to A. KOLMOGOROFF, A. KHINTCHINE in a paper (1934) returned to in the next section touches upon a continuous stationary process constructed by means of normal distribution functions. *Mutatis mutandis*, the same construction device will supply a discrete stationary process. In the sequel, this process will be termed the *normal process*. As a basis for illustrating the general stationary process, the normal process will prove very useful. In fact, in spite of the formal developments connected with the general normal process being of a simple structure, the normal process will, by proper specializations, be able to illustrate any type of stationary process mentioned in the previous section, and besides — as already advanced — the singular processes satisfying relations of type (83).

Before going into details concerning the normal stationary process, it will be convenient to introduce the concept of a general normal distribution in an enumerable set of variables.

Let an infinite quadratic form with real coefficients be given by

$$Q(X_1, X_2, \ldots) = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} \mu_{pq} \cdot X_p X_q. \tag{108}$$

In the following analysis, the variables X_i may take on any real values. Under such circumstances, the form Q will not always be convergent. If divergent, the form must be interpreted symbolically, viz. as the comprehension of all finite forms of type

$$Q_n(X_1, \ldots, X_n) = Q(X_1, X_2, \ldots, X_n, 0, 0, \ldots).$$

Thus prepared, let (A) $\mu_{pq} = \mu_{qp}$, and (B) any determinant $\Delta(n)$ of type

$$\Delta(n) = \begin{vmatrix} \mu_{11}, & \mu_{12}, & \ldots, & \mu_{1n} \\ \mu_{21}, & \mu_{22}, & \ldots, & \mu_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \mu_{n1}, & \mu_{n2}, & \ldots & \mu_{nn} \end{vmatrix}$$

be non-negative. Then, taking for $(m) = (m_1, m_2, \ldots)$ an arbitrary real sequence, a function $f(X_1, X_2, \ldots)$ in an enumerable set of variables $(X) = (X_1, X_2, \ldots)$ will be defined by

$$(109) \qquad f(X_1, X_2, \ldots) = e^{i \cdot \sum\limits_{p=1}^{\infty} m_p \cdot X_p - \frac{1}{2} \cdot Q(X_1, X_2, \ldots)}$$

As often as Q is divergent, this function must be interpreted symbolically in the same way as Q. Now, the function $f_n(X_1, \ldots, X_n)$ defined by

$$f_n(X_1, \ldots, X_n) = f(X_1, \ldots, X_n, 0, 0, \ldots) = e^{i \cdot \sum\limits_{p=1}^{n} m_p \cdot X_p - \frac{1}{2} \cdot Q_n(X_1, \ldots, X_n)}$$

is the characteristic function of a certain normal, n-dimensional aleatory variable, say $[\xi^{(1)}, \ldots, \xi^{(n)}]$, (see e. g. H. Cramér (1937) p. 109). We have $E[\xi^{(i)}] = m_i$, and $E[(\xi^{(i)} - m_i)(\xi^{(k)} - m_k)] = \mu_{ik}$. The form Q_n will thus be of the type (76).

In case $\Delta(n) \neq 0$, the variable $[\xi^{(1)}, \ldots, \xi^{(n)}]$ will possess an absolutely continuous distribution, and the density function, say $\varphi_n(u_1, \ldots, u_n)$, will be given by

$$\varphi_n(u_1, \ldots, u_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\Delta(n)}} \cdot e^{-\frac{1}{2} q_n(u_1, \ldots, u_n)}.$$

Here $q_n(u_1, \ldots, u_n)$ is defined by

$$q_n(u_1, \ldots, u_n) = \sum_{p=1}^{n} \sum_{q=1}^{n} \frac{\Delta_{pq}(n)}{\Delta(n)} \cdot (u_p - m_p)(u_q - m_q),$$

where $\Delta_{pq}(n)$ denotes the cofactor of $\Delta(n)$ belonging to the element μ_{pq}. Considering, on the other hand, the case $\Delta(n) = 0$, let it be assumed that $\Delta(n)$ is of rank $h < n$. Writing the relations of singularity on the form (69) after a suitable arrangement of the variables $\xi^{(i)}$, it follows from the previous analysis that the variable $[\zeta^{(1)}, \ldots, \zeta^{(n)}]$ defined by (72) will be composed of h, and only h, non-constant variables $\zeta^{(i)}$. On the other hand, for $i > h$ the variables $\zeta^{(i)}$ reduce to the constants m_i^* given by (75). Expressing the characteristic function of $[\zeta^{(1)}, \ldots, \zeta^{(n)}]$ in the variables Z_i, this will reduce to

$$(110) \qquad f_n^*(Z_1, \ldots, Z_n) = e^{i \cdot \sum\limits_{p=1}^{n} m_p^* \cdot Z_p - \frac{1}{2} Q_h^*(Z_1, \ldots, Z_h)}$$

Next, let us consider the infinite sequence $\xi = [\xi^{(1)}, \xi^{(2)}, \ldots]$. By construction, every finite sub-group $[\xi^{(i_1)}, \ldots, \xi^{(i_k)}]$ will possess a well-

defined probability distribution, and, further, satisfy all consistency relations of type (53—54). Thus ξ will constitute a random variable in an infinite number of dimensions. A variable ξ of this type will be termed *normal.* The function f will be called the *characteristic function* of ξ. This term is justified by the evident fact that the distribution of ξ is uniquely determined by f, and *vice versa.*

A necessary and sufficient condition that a normal distribution as defined by (109) may be taken to define a variable $\xi(t, t-1, \ldots.)$ connected with a stationary process $\{\xi(t)\}$ is that (a) m_p reduces to a constant, say m, and (b) $\mu_{p,q}$ is a function of $p-q$, say μ_{p-q}. In fact, taking $X_t, X_{t-1}, X_{t-2}, \ldots$ for variables in the characteristic function (109), the coefficients of X_{t-p} and of $X_{t-p} \cdot X_{t-q}$ will be independent of t when, and only when, the conditions (a) and (b) are satisfied simultaneously.

According to the condition (A) attached to (108), we have $\mu_{p-q} = \mu_{q-p} = \mu_{|p-q|}$. Further, disregarding the case of empty determinants $\Delta(n)$, the second condition, $\Delta(n) \geq 0$, is seen to imply $\mu_0 > 0$. Thus a set of real numbers $r_k = r_{-k}$ will be well-defined by putting

$$r_k = \mu_k/\mu_0.$$

In terms of these r_k, the conditions $\Delta(n) \geq 0$ will reduce to the inequalities (81).

According to the above, the general formula for the characteristic function of the variable $\xi(t, t-1, \ldots) = \xi[t]$ connected with a normal stationary process $\{\xi(t)\}$ is given by

$$(111) \qquad f(X_t, X_{t-1}, \ldots,) = e^{im \cdot \sum\limits_{p=0}^{\infty} X_{t-p} - \frac{\sigma^2}{2} \cdot \sum\limits_{p=0}^{\infty} \sum\limits_{q=0}^{\infty} r_{|p-q|} \cdot X_{t-p} X_{t-q}},$$

where (A) m, $\sigma > 0$, and r_k are real, and (B) the coefficients r_k satisfy the inequalities (81), viz. $\Delta(r, n) \geq 0$.

It is seen that the normal stationary process defined by (111) has for principal distribution function the normal distribution function

$$\Phi(u) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{u} e^{-\frac{1}{2}\sigma^2 (x-m)^2} dx,$$

and that the coefficients r_k appearing in (111) are nothing else than the autocorrelation coefficients of the process.

Thanks to theorem 2, the existence of a singular normal process satisfying (83) will be proved if the autocorrelation coefficients of any composed harmonic satisfying (30) are such that

$$\sum_{p=0}^{n} \sum_{q=0}^{n} r_{|p-q|} \cdot X_{t-p} X_{t-q} \geq 0 \tag{112}$$

for any n, and for any real sequence $(X) = (X_t, X_{t+1}, X_{t-2}, \ldots)$. For then, this form, and the quantity m obtained from (30), satisfy the conditions for introduction in the general formula (111) of the characteristic function of the variables $\xi(t, t-1, \ldots)$ defining the normal process, while the dispersion $\sigma > 0$ may by chosen freely.

The remaining proof involves no difficulty. Denoting by $x(t)$ an arbitrary composed harmonic satisfying (30), let a set of integers be given by $0 \leq i_1 < i_2 < \cdots < i_n$. Employing an argument used by A. Khintchine (1934), the obvious relations

$$\begin{aligned} (113) \quad 0 \leq \frac{1}{N} \cdot \sum_{s=1}^{N} [X_{t-i_1}(x(t-i_1+s)-m) + X_{t-i_2}(x(t-i_2+s)-m) + \cdots + \\ + X_{t-i_n}(x(t-i_n+s)-m)]^2 = \\ = \sum_{p=1}^{n} \sum_{q=1}^{n} X_{t-i_p} \cdot X_{t-i_q} \cdot \frac{1}{N} \sum_{s=1}^{N} (x(t-i_p+s) - m)(x(t-i_q+s) - m) \end{aligned}$$

will define a non-negative definite quadratic form in the variables X_{t-i_p}. After this observation, let $N \to \infty$. Then, disregarding a constant factor, the coefficient of $X_{t-i_p} \cdot X_{t-i_q}$ in the quadratic form will, by definition, tend to the autocorrelation coefficient $r_{|i_p-i_q|}$ which belongs to the composed harmonic considered (cf. (12)). Since the form will remain ≥ 0 also in the limit, and since the sequence i_p is arbitrary, the limit inequality implies (112).

Some circumstances connected with the singular processes merit particular attention. In the first place, assuming $L[\xi(t) - m] = 0$ as given by (77) to be the relation of singularity of the lowest possible order, the distribution of $\xi(t-1, \ldots, t-h)$ will be non-singular. In other words, if in a sample series

$$\xi_i(t_0 + k, t_0 + k - 1, \ldots, t_0 - h) = (\xi_{t_0+k}, \xi_{t_0+k-1}, \ldots, \xi_{t_0-h})$$

the values $\xi_{t_0-h}, \ldots, \xi_{t_0-2}$ are known, neither ξ_{t_0-1} nor $\xi_{t_0}, \xi_{t_0+1}, \ldots$ will be uniquely determined. On the other hand, if the sample series section $\xi_{t_0-h}, \ldots, \xi_{t_0-1}$ is known, then ξ_{t_0}, and, recurrently

$\xi_{t_0+1}, \ldots, \xi_{t_0+k}$ will, with probability one, be uniquely determined by means of $L[\xi_{t_0+t} - m] = 0$. Thus, ξ_{t_0+t} may be regarded as a solution of the difference equation (32) subject to the initial conditions $\xi_{t_0-1}, \ldots, \xi_{t_0-h}$. According to the theory of difference equations (cf. section 6), the periods p_k of the individual harmonics in ξ_{t_0+t} will be determined by the coefficients a_i in (77), and therefore be the same in all sample series connected with the process considered. On the other hand, the amplitudes C_k and the phases φ_k will be determined by the initial values. As mentioned above, these contain a random element, so the amplitudes and the phases of the individual harmonics constituting ξ_{t_0+t} will vary from one sample series $(\ldots, \xi_{t_0+k}, \xi_{t_0+k-1}, \ldots)$ to another. Of course, any expectation connected with the varying phases and amplitudes, e. g. $E[C_k^2]$, forms a characteristic of the process, and may, considering e. g. a normal process defined by (111), be expressed in terms of m, σ and r_k (cf. p. 73).

The above remarks apply, of course, to every singular process, and thus both to the periodic processes and to the singular normal processes. If the distribution of $\xi(t-1, \ldots, t-h)$ is absolutely continuous, which is always the case in the normal process, a more precise conclusion may be arrived at. In fact, let it in such case be assumed that not all of the individual harmonics in (17) would be present in ξ_{t_0+t}, i. e. that at least one harmonic would have a vanishing amplitude C_k. Then ξ_{t_0+t} must satisfy a difference equation of order $h-1$ having a general solution satisfying also (32). Since there are only h such equations at most, the sample series sections $(\xi_{t_0-1}, \ldots, \xi_{t_0-h})$ having the property assumed will form a set of Borel measure zero in the space of $\xi(t-1, \ldots, t-h)$. Keeping in mind the absolute continuity assumed, it follows that with probability one all individual harmonics really will present themselves when writing a sample series ξ_{t_0+t} on the form (17).

The investigations of E. Slutsky (see (1927) and, e. g., (1937)) and V. Romanovsky ((1932), (1933)) concerning the »*sinusoidal limit law*» fall under the theory of the stationary process, and present some parallelism with the previous analysis of the concept of singular process as introduced in section 14. Translating into the terminology of the present study, these authors investigate certain sequences of stationary processes, say $\{\beta^{(1)}(t)\}$, $\{\beta^{(2)}(t)\}, \ldots$ Denoting by $r_t^{(p)}$ the autocorrelation coefficients of the process $\{\beta^{(p)}(t)\}$, and by $x(t)$ a function (17) satisfying a linear relation $L[x(t) - m] = 0$,

the conditions imposed on the sequence $\{\beta^{(p)}(t)\}$ imply that $L[\gamma_t^{(p)}] \to 0$ as $p \to \infty$ (see V. ROMANOVSKY (1932), Theorem D). Representing by $(\beta_1^{(p)}, \ldots, \beta_n^{(p)})$ a section in a sample series of $\{\beta^{(p)}(t)\}$, and holding n fixed, the sinusoidal limit theorem asserts that, for sufficiently large values of p, the section considered will, with a probability as close to one as desired, approximate a function of type $x(t)$ with any prescribed accuracy.

A few reflections on the previous analysis will verify this theorem. Writing $m_p = E[\beta^{(p)}(t)]$, let an auxiliary set of processes $\{\xi^{(1)}(t)\}$, $\{\xi^{(2)}(t)\}, \ldots$ be defined by $\xi^{(p)}(t) = L[\beta^{(p)}(t) - m_p] = \Delta^{2s}\beta^{(p)}(t-s) + \cdots + h_s \cdot$ $\cdot [\beta^{(p)}(t) - m_p]$. It follows from the conditions imposed on $\gamma_t^{(p)}$ that the variables $\xi^{(p)}(t+k)$ will tend in probability to zero as $p \to \infty$. It remains to prove that the composite variable $[\xi^{(p)}(t), \xi^{(p)}(t+1), \ldots,$ $\xi^{(p)}(t+n)]$ will, for any fixed n, tend in probability to $(0, 0, \ldots, 0)$ as $p \to \infty$. This, however, follows at once from the BOOLE inequality (64).

By examples of the type $\{\beta^{(p)}(t)\} = \sum_{i=1}^{p} a_{pi} \cdot \{\alpha(t-i)\}$, E. SLUTSKY and V. ROMANOVSKY show that their theorems are not empty. While the added processes α used by SLUTSKY are all of the purely random type, ROMANOVSKY gives other examples as well. The recent paper of E. SLUTSKY (1937) already referred to contains references to certain related investigations, and illustrates in full detail the behaviour of model series of the processes $\{\beta^{(p)}(t)\}$ considered. In full agreement with the sinusoidal limit theorem, sections of moderate length in a model series approximate composed harmonics (17) of the proper type. The periods of the individual harmonics are the same in different sections, while the amplitudes and phases vary. These features are seen to be analogous to the properties of the singular processes proved above in connexion with the singular normal process. This parallelism is not accidental. In fact, an analysis of the sequences studied by E. SLUTSKY and V. ROMANOVSKY will show that these are convergent, and that the limit processes are singular. — In section 25 is given, i. a., a general device for the construction of sequences $\{\beta^{(p)}(t)\}$ ruled by the sinusoidal limit theorem.

17. The autocorrelation coefficients as FOURIER constants.

In a paper already referred to, A. KHINTCHINE (1934) studies what in a continuous statistical process with finite dispersion corresponds to the autocorrelation coefficients in a discrete process, viz. the function defined for any real u by

$$R(u) = E[(\xi(t) - E[\xi]) \cdot (\xi(t+u) - E[\xi])] / D^2(\xi),$$

where $\{\xi(t)\}$ represents the continuous stationary process considered. Terming $R(u)$ the *correlation function* of the process, KHINTCHINE gives, *i. a.*, a necessary and sufficient condition that a function $R(u)$ be the correlation function of a continuous stationary process. Slightly modifying the result, the condition is that there exists a distribution function, say $V(x)$, such that $V(0) = 0$, and

$$R(u) = \int_0^\infty \cos ux \cdot d\,V(x). \tag{114}$$

The inversion formula (see e. g. H. CRAMÉR (1937), Theorem 9)

$$V(x) = \frac{2}{\pi} \int_0^\infty R(u) \frac{\sin u x}{u} d\,u$$

shows that the function $V(x)$ is uniquely determined by $R(u)$.

We shall first give a corresponding theorem on the discrete stationary process. It should be observed that the same theorem holds for the generalized process (cf. section 11).

Theorem 5. Let r_k ($k = 0, \pm 1, \pm 2, \ldots$) be an arbitrary sequence of constants. A necessary and sufficient condition that there exists a discrete stationary process with the r_k's for autocorrelation coefficients is that the r_k's are the FOURIER *coefficients of a non-decreasing function, say $W(x)$, such that $W(0) = 0$; $W(\pi) = \pi$,*

$$r_k = \frac{1}{\pi} \int_0^\pi \cos kx \cdot d\,W(x). \tag{115}$$

Before passing to the proof we give the following inversion formula involving a converging sum (see e. g. F. HAUSDORFF (1923) p. 245)

$$W(x) = x + 2 \cdot \sum_{k=1}^{\infty} \frac{r_k}{k} \sin kx. \tag{116}$$

With suitable agreements as to points of discontinuity, the function $W(x)$ thus will be uniquely determined by the autocorrelation coefficients. In the sequel, $W(x)$ will be called the *generating function* of the autocorrelation coefficients r_k.

The proof will use some facts concerning definite quadratic forms, facts parallel to the properties of definite functions used by A. KHINTCHINE in the continuous case. In other respects the proofs are coincident.

For a verification of the necessity of the condition given, let $\{\xi(t)\}$ be an arbitrary discrete stationary process with finite dispersion, say σ. Put $E[\xi(t)] = m$, let r_k represent the autocorrelation coefficients of the process $\{\xi(t)\}$, and consider the quadratic form

$$\sum_{p=0}^{n} \sum_{q=0}^{n} r_{|p-q|} \cdot X_p X_q, \qquad n = 0, 1, 2, \ldots \tag{117}$$

For any real sequence $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ we have (cf. (113))

$$0 \leq \frac{1}{\sigma^2} E[\sum_{p=1}^{n} X_{t_p} [\xi(t\ ) - m]]^2 =$$

$$= \frac{1}{\sigma^2} \cdot \sum_{p=1}^{n} \sum_{q=1}^{n} E[X_{t_p} X_{t_q} [\xi(t_p) - m][\xi(t_q) - m]] = \sum_{p=1}^{n} \sum_{q=1}^{n} r_{|t_p - t_q|} \cdot X_{t_p} X_{t_q}.$$

This relation implies that the forms (117) are non-negative definite. Next, according to a theorem of G. HERGLOTZ* (1911), this statement is equivalent to saying that the following system of trigonometrical moments,

$$\frac{1}{2\pi} \int_0^{2\pi} \cos kx \cdot dW(x) = r_k; \qquad \frac{1}{2\pi} \int_0^{2\pi} \sin kx \cdot dW(x) = 0, \tag{118}$$

has a non-decreasing solution $W(x)$ with $W(0) = 0$.

The inversion of (118) gives exactly (116). The required relation (115) follows directly from (116), and since (116) further gives

* Prof. T. CARLEMAN has kindly informed me that the formulae (118) may be obtained directly from the HILBERT representation of a definite quadratic form of general type.

$$W(2\pi - x) = 2\pi - x - 2 \cdot \sum_{k=1}^{\infty} \frac{r_k}{k} \sin kx,$$

we obtain $W(x) = 2\pi - W(2\pi - x)$, and, finally, $W(\pi) = \pi$.

On the other hand, if r_k is given by (115), the HERGLOTZ theorem asserts that the forms (117) are non-negative. Thus, the relation (112) holds, and there exists a normal process with the given r_k-values for autocorrelation coefficients.

The theorem proved above permits of some general conclusions concerning the autocorrelation coefficients of a discrete stationary process. Considering the coefficients r_k for large k-values, their behaviour will depend on the continuity structure of the generating function $W(x)$. In order to study the situation in some detail, let

$$W(x) = a^{(1)} \cdot W^{(1)}(x) + a^{(2)} \cdot W^{(2)}(x) + a^{(3)} \cdot W^{(3)}(x)$$

stand for the well-known (cf. e. g. H. CRAMÉR (1928), p. 59) representation of $W(x)$ as the sum of three uniquely determined, non-decreasing functions with $W^{(1)}(0) = W^{(2)}(0) = W^{(3)}(0) = 0$, $W^{(1)}(\pi) = W^{(2)}(\pi) = W^{(3)}(\pi) = \pi$, $a^{(1)} + a^{(2)} + a^{(3)} = 1$, $a^{(i)} \geq 0$, and

1) $a^{(1)} \cdot W^{(1)}(x) = \int_0^{\pi} \frac{dW(y)}{dy} \cdot dy,$

2) $a^{(2)} \cdot W^{(2)}(x)$, the *saltus* function, is equal to the sum of saltuses of $W(x)$ at all the points of discontinuity which are less than or equal to x; writing λ_ν for the saltus points of $W(x)$, and $c_\nu^2 \cdot \pi/2$ for the corresponding saltuses, then

$$\frac{a^{(2)}}{\pi} \cdot W^{(2)}(x) = \sum_{\lambda_\nu \leq x} c_\nu^2/2,$$

3) $a^{(3)} \cdot W^{(3)}(x)$, the *singular* function, is a continuous function which has almost everywhere a derivative equal to zero.

Thus prepared, we put

$$r_k^{(i)} = \frac{1}{\pi} \int_0^{\pi} \cos kx \cdot dW^{(i)}(x), \qquad i = 1, 2, 3,$$

and obtain

(119) $$r_k = a^{(1)} \cdot r_k^{(1)} + a^{(2)} \cdot r_k^{(2)} + a^{(3)} \cdot r_k^{(3)}.$$

The components $r_k^{(i)}$ thus uniquely determined by the r_k-sequence

via its generating function (cf. (116)) are of entirely different character.

As to $r_k^{(1)}$ we have (see e. g. H. C. CARSLAW (1930), p. 271)

$$a^{(1)} \cdot r_k^{(1)} = \frac{1}{\pi} \int_0^\pi \frac{d\,W(x)}{d\,x} \cos k\,x\,d\,x \to 0 \text{ as } k \to \infty.$$

The component $r_k^{(2)}$ is given by

$$a^{(2)} \cdot r_k^{(2)} = \frac{1}{2} \cdot \sum_{i=1}^{\infty} c_i^2 \cdot \cos \lambda_i\,k. \tag{120}$$

It is seen that $r_k^{(2)}$ is an almost periodic function. Again referring to (19), we conclude that arbitrarily large k-values exist for which $r_k^{(2)}$ approximates

$$r_0^{(2)} = \frac{1}{\pi} W^{(2)}(\pi) = \frac{1}{2\,a^{(2)}} \cdot \sum_{i=1}^{\infty} c_i^2.$$

The singular component $r_k^{(3)}$ permits of no unconditioned statement as to its behaviour for large k-values.

In order to arrive at a criterion of the structure of $W(x)$, let it be assumed that $\sum_{k=1}^{\infty} |r_k|$ is convergent. It follows from (116) that in such a case $W(x)$ for all x has a derivative $W'(x)$, that this will be obtained by summing the derivatives of the terms in the right member of (116), and that $W'(x)$ will be bounded. We thus obtain the following corollary to theorem 5.

Corollary. Let $\{\xi(t)\}$ be a stationary process with autocorrelation coefficients r_k such that $\sum |r_k|$ is convergent. Then $W(x)$ will be absolutely continuous,

$$W(x) = W^{(1)}(x).$$

The derivative $W'(x)$ is bounded in modulus, and given by

$$W'(x) = \sum_{k=-\infty}^{\infty} r_k \cos k\,x, \qquad 0 \leq x < \pi. \tag{121}$$

As a first application of the above analysis, we shall touch upon some problems concerning the relation between continuous and discrete stationary processes. The question to be put corresponds to a problem dealt with by G. ELFVING (1937) in a study on MARKOFF chains.

A continuous stationary process, say $\{\xi^{(c)}(t)\}$, gives a hypothetical scheme for the probability relations in any time points by means of a set $\{F^{(c)}\}$ of distribution functions $F(t_1, \ldots, t_n; u_1, \ldots, u_n)$ satisfying (53)—(55) and referring to quite arbitrary time points $t_1, \ldots, t_n$. Among these distribution functions, let those referring to integral time points be represented by $\{F^{(d)}\}$. It is plain that the set $\{F^{(d)}\}$ thus obtained will define a discrete stationary process, say $\{\xi^{(d)}(t)\}$. The situation may be described by saying that the hypothesis $\{\xi^{(c)}(t)\}$ is consistent with the hypothesis $\{\xi^{(d)}(t)\}$.

Marking the symbols referring to consistent processes $\{\xi^{(c)}\}$ and $\{\xi^{(d)}\}$ by (c) and (d) respectively, let $\{\xi^{(c)}\}$ have a finite dispersion. Then, evidently, we have for any integral k (cf. (114))

$$r_k^{(d)} = R^{(c)}(k).$$

Further, for any x in the interval $(0, \pi)$, we have

$$W(x) = \pi \cdot \sum_{n=0}^{\infty} [V(n \cdot 2\pi + x) - V(n \cdot 2\pi - x)]. \tag{122}$$

In fact, inserting the right member of (122) in (115), and paying regard to (114), we obtain by elementary transformations for integral k-values

$$r_k^{(d)} = \int_0^{\pi} \cos kx \cdot \sum_{n=0}^{\infty} d[V(n \cdot 2\pi + x) - V(n \cdot 2\pi - x)] = R^{(c)}(k).$$

We see that (122) illustrates the ambiguity about periods in the discrete case (cf. p. 15). A harmonic with period $p < 2$ will in the continuous case correspond to an increase in $V(\lambda)$ for $\lambda > \pi$. In the discrete case, such a harmonic being read off only for integral time points, its values vill coincide with those of a certain harmonic with period $p > 2$ and frequency $\lambda < \pi$.

On the other hand, let a discrete process $\{\xi^{(d)}(t)\}$ be defined by a set $\{F^{(d)}\}$ of distribution functions referring to integral time points. Our question is whether there exists an »interpolating» continuous process $\{\xi^{(c)}(t)\}$ defined by a set $\{F^{(c)}\}$ where the distribution functions referring to integral time points are identical with the given set $\{F^{(d)}\}$. Studying in the first place the autocorrelation coefficients $r_k^{(d)} = \frac{1}{\pi} \int_0^{\pi} \cos kx \, dW(x)$, we seek a continuous stationary process $\{\xi^{(c)}\}$ with correlation function $R^{(c)}(u)$ such that, for

integral k-values, $R^{(c)}(k) = r_k^{(d)}$. Taking $V(x) = \frac{1}{\pi} W(x)$ for $0 \leq x \leq \pi$, and $V(x) = 1$ for $x \geq \pi$, such a function is evidently yielded by

$$R^{(c)}(u) = \int_0^\infty \cos ux \cdot dV(x). \tag{123}$$

It should be observed that (123) remains unchanged for integral u-values when substituting, for instance, $V(x - n \cdot 2\pi)$ for $V(x)$, letting n denote a positive integer. Thus, the interpolating function $R^{(c)}(u)$ will by no means be uniquely determined by the autocorrelation coefficients $r_k^{(d)}$ given. This indeterminateness is, of course, analogous to the circumstance mentioned in section 4 that there exists an infinite number of simple harmonics all of which pass through all of the values $x(t_n)$ taken on in equidistant points t_n by a simple harmonic $x(t)$.

In case the process $\{\xi^{(d)}\}$ considered is normal, the Khintchine-Kolmogoroff device for constructing a normal continuous process may be applied on the basis of an interpolating $R^{(c)}(u)$ as given, for instance, by (123). Furnishing the normal continuous process with the same mean and the same dispersion as $\{\xi^{(d)}\}$, the resulting process will possess the property desired, i. e. give rise to the primary discrete process $\{\xi^{(d)}\}$ when considering the probability relations in integral time points.

Illustration. Taking $V(x) = 1 - e^{-x}$, formula (114) gives

$$R^{(c)}(u) = 1/(1 + u^2).$$

The sum (122) is readily computed, and gives

$$W(x) = \pi \left[1 - \frac{e^{-x} - e^{x-2\pi}}{1 - e^{-2\pi}}\right]; \qquad 0 \leq x \leq \pi. \tag{124}$$

Insertion of this expression in (115) gives without difficulty

$$r_k^{(d)} = \frac{1}{1 + k^2} + \frac{2k \sin \pi k}{(e^\pi - e^{-\pi})(1 + k^2)}. \tag{125}$$

In full agreement with the general results, we find that for integral k-values $r_k^{(d)} = R^{(c)}(k)$.

On the other hand, in case a discrete process is given, and has $W(x)$ as defined by (124) for generating function, one of the interpolating correlation functions will be obtained by taking $R^{(c)}(u) = r_u^{(d)}$, where $r_k^{(d)}$ for $-\infty < k < \infty$ is given by (125).

The corresponding function $V(x)$ is, of course, equal to $\frac{1}{\pi} W(x)$ in $(0, \pi)$ and equal to 1 in (π, ∞).

The second application of formula (115) is concerned with the periodogram method for graduating a sample series section connected with a stationary process. Only discrete processes will be considered; the arguments also hold, however, in the continuous case.

First, an investigation will be made as to whether a SCHUSTER periodogram analysis of a sample series section, say $\xi_{t-1}, \ldots, \xi_{t-n}$, may be expected to be effective, in particular for large n-values. It will turn out that if the generating function of the process has a non-vanishing saltus component, the periodogram analysis will give positive results, viz. in the sense that the expectance of certain well-defined periodogram ordinates will be positive. On the other hand, in case the corollary of theorem 5 applies, a periodogram analysis will prove resultless.

Let $(\xi_{t-1}, \xi_{t-2}, \ldots)$ stand for a sample series belonging to a discrete stationary process $\{\xi(t)\}$ with dispersion σ, mean m, and generating function $W(x)$.

Applying the classical SCHUSTER periodogram analysis to the sample series section $(\xi_{t-1}, \xi_{t-2}, \ldots, \xi_{t-n})$, let the resulting periodogram functions be denoted by (see (26) and (27))

$$A(n, \lambda) = \frac{2}{n} \sum_{p=1}^{n} (\xi_{t_0+p} - m) \cos \lambda p; \qquad B(n, \lambda) = \frac{2}{n} \sum_{q=1}^{n} (\xi_{t_0+q} - m) \sin \lambda q;$$

$$C^2(n, \lambda) = A^2(n, \lambda) + B^2(n, \lambda), \tag{126}$$

where $t_0 = t - n - 1$. Studying in the first place the expression

$$E = \lim_{n \to \infty} E[C^2(n, \lambda)],$$

elementary transformations yield

$$\begin{aligned}(127)\quad & E[C^2(n, \lambda)] = \\ &= \frac{4}{n^2} \sum_{p=1}^{n} \sum_{q=1}^{n} (\cos \lambda p \cdot \cos \lambda q + \sin \lambda p \cdot \sin \lambda q) E[(\xi(t_0+p) - m)(\xi(t_0+q) - m)] = \\ &= \frac{4\sigma^2}{n^2} \sum_{p=1}^{n} \sum_{q=1}^{n} r_{|p-q|} \cos \lambda (p-q) = \frac{4\sigma^2}{n} \sum_{k=-n+1}^{n-1} \frac{n - |k|}{n} r_k \cos \lambda k = \\ &= \frac{4\sigma^2}{n} \left[1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) r_k \cos \lambda k\right].\end{aligned}$$

In case the process $\{\xi(t)\}$ is non-autocorrelated, we have $r_k = 0$ for $k > 0$. The above formula then reduces to the SCHUSTER formula (37).

In a study on sampling problems in intercorrelated series, E. SLUTSKY (1934) investigates, i. a., expectations of type $E[C^2(n, \lambda)]$ for λ-values equalling multiples of $2\pi/n$, i. e. expectations connected with the FOURIER coefficients of a sample series section $(\xi_{t-1}, \ldots, \xi_{t-n})$ (cf. (25)). Under certain restrictive conditions concerning the process considered, SLUTSKY gives the relations (127).

J. BARTELS (1935) seems to be the first to have deduced the relations (127) without setting restrictive conditions on the process analysed.[6]

In order to avoid discussions which might obscure the point of the analysis, we shall now introduce a restriction concerning the generating function of the autocorrelation coefficients, viz. that $a^{(3)} = 0$, and that $\Sigma \, |r_k^{(1)}|$ is convergent. According to (121), the latter assumption implies that $d\,W^{(1)}(\lambda)/d\,\lambda$ is finite for all λ.

It is seen that the limit expectation E may be split up into portions corresponding to (119), say $a^{(1)} \cdot E^{(1)}$, $a^{(2)} \cdot E^{(2)}$ and $a^{(3)} \cdot E^{(3)}$. By the simplifying hypotheses made, we have $a^{(3)} = 0$.

Considering the relations (127), we observe that an elementary transformation gives

$$(128) \qquad 1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) r_k \cos \lambda k = \frac{1}{n} \cdot \sum_{s=0}^{n-1} \sum_{k=-s}^{s} r_k \cos \lambda k,$$

and that the right member is a CESÀRO mean, viz. the arithmetical mean of the first n partial sums $\sum_{-s}^{s} r_k \cos \lambda k$ of the series appearing in (121). After substitution of $r_k^{(1)}$ for r_k, the expression (128) thus will tend to $d\,W^{(1)}(\lambda)/d\,\lambda$ as $n \to \infty$. As by hypothesis this derivative is bounded in modulus, we conclude that for all λ in $(0, \pi)$

$$(129) \qquad E^{(1)} = \frac{4\,\sigma^2}{n} \cdot \frac{1}{a^{(1)}} \cdot \frac{d\,W^{(1)}(\lambda)}{d\,\lambda} + o(1/n) \to 0 \text{ as } n \to \infty.$$

Paying regard to (120), a similar argument gives

$$(130) \qquad \begin{cases} E^{(2)} = \lim\limits_{n \to \infty} \dfrac{4\,\sigma^2}{n} \left[1 + 2 \sum\limits_{k=1}^{n-1} \left(1 - \dfrac{k}{n}\right) r_k^{(2)} \cos \lambda k\right] = 0 \text{ for } \lambda \neq \lambda_\nu, \\ E^{(2)} = c_\nu^2 \cdot \sigma^2 / a^{(2)} \text{ for } \lambda = \lambda_\nu < \pi. \end{cases}$$

Thus, while an ordinate $C^2(n, \lambda_\nu)$ in the periodogram of a sample series may vary from one series to another, its expectation is by a simple limit relation connected with the saltus, $\frac{\pi}{2}\,c_\nu^2$, in $\lambda = \lambda_\nu$ of the generating function $W(\lambda)$ of the autocorrelation coefficients.

It is seen from (129)—(130) that, under the assumptions made, it is only if $a^{(2)} > 0$ that a periodogram analysis of a sample series will be fruitful. Assuming that there are s discontinuities in $W^{(2)}(\lambda)$, the analysis will result in a composed harmonic,

$$x_n(t-k) = \sum_{\nu=1}^{s} A(n, \lambda_\nu) \cos(n+1-k)\lambda_\nu + \sum_{\nu=1}^{s} B(n, \lambda_\nu) \sin(n+1-k)\lambda_\nu,$$

approximating the section $(\xi_{t-1}, \ldots, \xi_{t-n})$ analysed, and with coefficients depending on the sample series considered and satisfying

$$\lim_{n \to \infty} \frac{1}{\sigma^2} E[A^2(n, \lambda_\nu) + B^2(n, \lambda_\nu)] = c_\nu^2.$$

A standard measure of the deviation $\xi_{t-k} - x_n(t-k)$ is given by the expectation E defined by

$$E = E\left[\frac{1}{n} \cdot \sum_{k=1}^{n} [\xi(t-k) - x_n(t-k)]^2\right] \cdot$$

Disregarding terms of order $1/n$, the coefficients A and B in $x_n(t-k)$ will make E a minimum (cf. (43)),

$$E \cong (1 - \frac{1}{2} c_1^2 - \cdots - \frac{1}{2} c_s^2) \cdot \sigma^2 = \left[1 - \frac{a^{(2)}}{\pi} \cdot W^{(2)}(\pi)\right] \cdot \sigma^2.$$

This relation shows clearly the scope of the periodogram analysis. In fact, in the special case $W(x) \equiv W^{(2)}(x)$, the approximating function $x_n(t-k)$ will, for sufficiently large n-values, yield a fit as close as desired.

On the other hand, it follows from the previous analysis that in case $\Sigma |r_k|$ is convergent, then $W(x) \equiv W^{(1)}(x)$, and $d\,W(x) / d\,x$ is bounded. In this case, a periodogram analysis of the section $(\xi_{t-1}, \xi_{t-2}, \ldots)$ will be resultless (cf. (129)).[7] Processes of this type have to be attacked by the use of entirely different methods. Now, a method at once suggesting itself is that of linear regression analysis. For instance, approximating $\xi(t)$ linearly by means of $\xi(t-1)$ we obtain $\xi(t) = r_1 \cdot \xi(t-1) + \eta(t)$, denoting by $\eta(t)$ a residual variable with variance $\sigma^2 \cdot (1 - r_1^2)$. Otherwise expressed, $r_1 \cdot \xi_{t-1}$ yields an effective prognosis of ξ_t with a squared dispersion equalling $E[[\xi_t - r_1 \cdot \xi_{t-1}]^2] = \sigma^2 \cdot (1 - r_1^2)$. This simple instance is sufficient for exemplifying that this method is of quite another

type than the periodogram analysis. The next stage of the present study will be to follow up this line of research; this is done in section 19. The coming section is reserved for a preparatory survey.

For illustrations of periodogram analysis of model series, reference is given to section 25. For the present we only remark that the general formula (127) gives

$$\frac{1}{\pi}\int_0^\pi E[C^2(n, \lambda)]\, d\lambda = \frac{4\sigma^2}{n}. \tag{131}$$

Comparing with (37), the expectation of the ordinates in a periodogram is seen to be a function of λ with mean value equalling the constant (in respect of λ) expectance in the case of a purely random process.

It is rather interesting to compare the relations (131) and (130). It is seen that (131) also holds in the case of a process of hidden periodicities, and that in such a process the ordinates $E[C^2(n; \lambda_\nu)]$ will not tend to zero as $n \to \infty$. We conclude that the rise to a maximum in $E[C^2(n; \lambda)]$ will be very rapid, and that the corresponding peaks in the periodogram will be very thin, with a breadth of an order of magnitude not surpassing $1/n$.

On the other hand, if the corollary to theorem 5 applies, the expectation $E[C^2(n; \lambda)]$ evidently tends to zero uniformly in λ as $n \to \infty$,

$$E[C^2(n; \lambda)] = \frac{4W'(\lambda)}{n}\cdot\sigma^2 + o(1/n).$$

In full agreement with (121) we conclude from (131) that in the present case $\frac{1}{\pi}\int_0^\pi W'(\lambda)\, d\lambda = 1.$

18. On linear approximation in a space of random variables.

The next section presenting a generalization of the regression analysis of ordinary random variables to the general discrete stationary process, the present section is reserved for an interpretation of the ordinary regression analysis as a linear approximation in a metrical space. The interpretation applies, of course, to statistical as well as to aleatory random variables.

When dealing in the following with a set of random variables, say ξ, we shall tacitly assume that any finite sub-group, say $[\xi^{(1)}, \ldots,$

$\xi^{(n)}]$, forms a well-defined multi-dimensional variable, say with distribution function $F(u_1, \ldots, u_n)$, and that the functions F are consistent (cf. p. 32 f.). As observed by M. FRÉCHET ((1937), p. 205 f.), a set of one-dimensional random variables with finite dispersion can be made metrical by defining the distance between two variables in the set, say $\xi^{(i)}$ and $\xi^{(k)}$, as the dispersion of the difference variable $\xi^{(i)}-\xi^{(k)}$,

$$(132) \qquad |\xi^{(i)} - \xi^{(k)}| = D(\xi^{(i)} - \xi^{(k)}).$$

This fact depends on the triangular inequality

$$(133) \qquad D(\xi^{(1)} - \xi^{(2)}) + D(\xi^{(1)} - \xi^{(3)}) \geq D(\xi^{(2)} - \xi^{(3)}),$$

which by elementary transformations reduces to the inequality of SCHWARZ.

Adopting the distance definition (132), next let a stand for a real number, and let ξ and η be two random variables. Then we have

$$(134) \qquad |\xi - a \cdot \eta|^2 = D^2(\xi) - 2\,a \cdot r(\xi, \eta) \cdot D(\xi) \cdot D(\eta) + a^2 D^2(\eta).$$

This squared distance will reach a minimum equalling

$$(135) \qquad [1 - r^2(\xi, \eta)] \cdot D^2(\xi)$$

when a equals the regression coefficient of ξ on η, i. e. when

$$(136) \qquad a = r(\xi, \eta) \cdot D(\xi)/D(\eta).$$

This regression coefficient is linear in respect to ξ. In fact,

$$(137) \qquad r(\xi + \zeta, \eta) \cdot D(\xi + \zeta) = \frac{E[(\xi - E[\xi] + \zeta - E[\zeta])(\eta - E[\eta])]}{D(\eta)} =$$

$$= \frac{E[(\xi - E[\xi])(\eta - E[\eta])] + E[(\zeta - E[\zeta])(\eta - E[\eta])]}{D(\eta)} =$$

$$= r(\xi, \eta) \cdot D(\xi) + r(\zeta, \eta) \cdot D(\zeta).$$

Using these properties of regression coefficients, the multiple regression theory founded and developed by the English statistical school (see G. U. YULE and M. G. KENDALL (1937), p. 511, for references) can be interpreted as a particular branch of the theory of approximations in general linear spaces. In the general terminology, uncorrelated variables $\xi^{(i)}$ and $\xi^{(k)}$ should be called orthogonal elements in the space to which they belong. For later application we shall record some general approximation formulae in terms of determinants. The verification being in detail parallel to the GRAM-SCHMIDT orthogonalization of vectors in an infinite number of dimensions, reference is again made to G. KOWALEWSKI ((1909), § 175).[8]

Let an n-dimensional variable $[\xi^{(1)}, \ldots, \xi^{(n)}]$ formed by variables $\xi^{(i)}$ with finite dispersion be linearly non-singular (see p. 41 f.). Considering an arbitrary variable $\xi^{(0)}$ with finite dispersion, and writing $m_i = E[\xi^{(i)}]$; $\mu_{ik} = E[(\xi^{(i)} - m_i)(\xi^{(k)} - m_k)]$, there exists a well-defined sequence of coefficients a_{in} minimizing the dispersion of the variable defined for arbitrary a_{in}'s by

$$\xi^{(0)} - m_0 - a_{1n} \cdot (\xi^{(1)} - m_1) - \cdots - a_{nn} \cdot (\xi^{(n)} - m_n). \tag{138}$$

Terming *residual*, and denoting by $\eta^{(n)}$ the variable (138) formed by the minimizing coefficients a_{in}, we have $E[\eta^{(n)}] = 0$, and

$$\eta^{(n)} = \frac{\begin{vmatrix} \mu_{11}, \ldots, \mu_{1n}, \xi^{(1)} - m_1 \\ \ldots\ldots \\ \mu_{n1}, \ldots, \mu_{nn}, \xi^{(n)} - m_n \\ \mu_{01}, \ldots, \mu_{0n}, \xi^{(0)} - m_0 \end{vmatrix}}{\begin{vmatrix} \mu_{11}, \ldots, \mu_{1n} \\ \ldots \\ \mu_{n1}, \ldots, \mu_{nn} \end{vmatrix}}; \quad D^2(\eta^{(n)}) = \frac{\begin{vmatrix} \mu_{11}, \ldots, \mu_{1n}, \mu_{10} \\ \ldots\ldots \\ \mu_{n1}, \ldots, \mu_{nn}, \mu_{n0} \\ \mu_{01}, \ldots, \mu_{0n}, \mu_{00} \end{vmatrix}}{\begin{vmatrix} \mu_{11}, \ldots, \mu_{1n} \\ \ldots \\ \mu_{n1}, \ldots, \mu_{nn} \end{vmatrix}}. \tag{139}$$

A set of formulae equivalent to (139), and involving an auxiliary set of one-dimensional aleatory variables $\zeta^{(i)}$, is given by

$$\eta^{(n)} = \xi^{(0)} - m_0 - c_1 \cdot \zeta^{(1)} - c_2 \cdot \zeta^{(2)} - \cdots - c_n \cdot \zeta^{(n)}, \tag{140}$$

$$D^2(\eta^{(n)}) = D^2(\xi^{(0)}) - c_1^2 - c_2^2 - \cdots - c_n^2. \tag{141}$$

For the auxiliary variables we have (cf. G. Kowalewski (1909) p. 426 and T. Lindblad (1937)),

$$\zeta^{(1)} = \frac{\xi^{(1)} - m_1}{\sqrt{1 \cdot \mu_{11}}}; \quad \zeta^{(2)} = \frac{\begin{vmatrix} \mu_{11}, \xi^{(1)} - m_1 \\ \mu_{21}, \xi^{(2)} - m_2 \end{vmatrix}}{\sqrt{\mu_{11} \cdot \begin{vmatrix} \mu_{11}, \mu_{12} \\ \mu_{21}, \mu_{22} \end{vmatrix}}}; \tag{142}$$

$$\zeta^{(3)} = \frac{\begin{vmatrix} \mu_{11}, \mu_{12}, \xi^{(1)} - m_1 \\ \mu_{21}, \mu_{22}, \xi^{(2)} - m_2 \\ \mu_{31}, \mu_{32}, \xi^{(3)} - m_3 \end{vmatrix}}{\sqrt{\begin{vmatrix} \mu_{11}, \mu_{12} \\ \mu_{21}, \mu_{22} \end{vmatrix} \cdot \begin{vmatrix} \mu_{11}, \mu_{12}, \mu_{13} \\ \mu_{21}, \mu_{22}, \mu_{23} \\ \mu_{31}, \mu_{32}, \mu_{33} \end{vmatrix}}};$$

. . . .

These variables are *standardized*, i. e.

(143) $$E[\zeta^{(i)}]=0; \quad D[\zeta^{(i)}]=1.$$

They are further mutually non-correlated,

(144) $$r(\zeta^{(i)}, \zeta^{(k)}) = E[\zeta^{(i)} \cdot \zeta^{(k)}] = 0 \text{ for } i \neq k.$$

and it is thanks to this relation that the coefficients c_i in (140) are independent of n,

(145) $$c_i = r(\xi^{(0)}, \zeta^{(i)}) \cdot D(\xi^{(0)}) = E[(\xi^{(0)} - m_0) \cdot \zeta^{(i)}].$$

On the other hand, in case $[\xi^{(1)}, .., \xi^{(n)}]$ is singular, say of rank h [i. e., by relations of the type (69)], the cofffficients a_{in} in the residual variable $\eta^{(n)}$ will not be uniquely determined. In fact, the addition of an arbitrary linear combination of the vanishing sums (69) will not change the variable (138). In full agreement herewith, the expressions (139) become indeterminate when $[\xi^{(1)}, .., \xi^{(n)}]$ is taken to be singular.

Whether or not $[\xi^{(1)}, .., \xi^{(n)}]$ be singular, the residual variable $\eta^{(n)}$ will be uncorrelated with every $\xi^{(i)}$. In fact, formulae (134)—(136) imply that otherwise the variance $D^2(\eta^{(n)})$ could be brought down by subtraction in $\eta^{(n)}$ of the non-vanishing variable

$$r(\eta^{(n)}, \xi^{(i)}) \frac{D(\eta^{(n)})}{D(\xi^{(i)})} \cdot [\xi^{(i)} - m_i].$$

Considering the residuals $\eta^{(1)}, \eta^{(2)}, \ldots, \eta^{(n)}$, it follows from (141) that the variances $D^2(\eta^{(1)}), D^2(\eta^{(2)}), \ldots, D^2(\eta^{(n)})$ form a non-increasing sequence. It is of central importance that, when $D^2(\eta^{(i)}) > 0$, we shall have $D^2(\eta^{(i+1)}) = D^2(\eta^{(i)})$ if, and only if, one of the following two cases is present:[9]

(A) $[\xi^{(1)}, .., \xi^{(i+1)}]$ is singular by means of a relation of type

(146) $$p_1 \cdot (\xi^{(1)} - m_1) + \cdots + p_i \cdot (\xi^{(i)} - m_i) = \xi^{(i+1)} - m_{i+1}$$

(B) No relation of type (146) exists, but $\xi^{(i+1)}$ is uncorrelated with $\eta^{(i)}$.

It follows, i. a., that if $[\xi^{(1)} .., \xi^{(n)}]$ is singular of rank h, the variables $\xi^{(i)}$ may be arranged in an order such that the coefficients a_{pq} in the residuals $\eta^{(1)}, \eta^{(2)}, \ldots, \eta^{(h)}$, and only in these, are uniquely determined, while

$$(147)\qquad D(\xi^{(0)}) \geq D(\eta^{(1)}) \geq D(\eta^{(2)}) \geq \cdots \geq D(\eta^{(h)}) = \\ = D(\eta^{(h+1)}) = \cdots = D(\eta^{(n)}) \geq 0.$$

In other words, the first h variables $\xi^{(i)}$ alone will be able to bring down the residual dispersion to its minimum value.

The following determinant expression for a general partial correlation coefficient in the notation of G. U. YULE (see G. U. YULE and M. G. KENDALL (1937), p. 269) will attach the above system of formulae to the familiar theory of multiple correlation. The formula is valid in case none of the variables $[\xi^{(0)}, \xi^{(2)}, \xi^{(3)}, \ldots, \xi^{(n)}]$ and $[\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(n)}]$ is singular.

$$r_{01,23\ldots n} = \frac{\begin{vmatrix} \mu_{01}, & \mu_{02}, & \ldots, & \mu_{0n} \\ \mu_{21}, & \mu_{22}, & \ldots, & \mu_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \mu_{n1}, & \mu_{n2}, & \ldots, & \mu_{nn} \end{vmatrix}}{\sqrt{\begin{vmatrix} \mu_{11}, & \mu_{12}, & \ldots, & \mu_{1n} \\ \mu_{21}, & \mu_{22}, & \ldots, & \mu_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \mu_{n1}, & \mu_{n2}, & \ldots, & \mu_{nn} \end{vmatrix} \cdot \begin{vmatrix} \mu_{00}, & \mu_{02}, & \ldots, & \mu_{0n} \\ \mu_{20}, & \mu_{22}, & \ldots, & \mu_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \mu_{n0}, & \mu_{n2}, & \ldots, & \mu_{nn} \end{vmatrix}}}$$

A proof from n to $n+1$ will verify

$$D^2(\eta^{(n)}) = \mu_{00} \cdot (1 - r_{01}^2)(1 - r_{02,1}^2) \cdots (1 - r_{0n,12\ldots(n-1)}^2),$$

a formula which shows, i. a., that also the partial correlation coefficients lie in the interval $(-1, 1)$.

For later application, we record that

$$(148)\qquad a_{1n} = r_{01,23\ldots n} \cdot \frac{D_{0,23\ldots n}}{D_{1,23\ldots n}},$$

and analogous formulae hold for the remaining coefficients a_{in} (cf. G. U. YULE and M. G. KENDALL (1937), p. 266). Here $D_{i,23\ldots n}$ for $i = 0$ or 1 represents the dispersion of the residual variable obtained when approximating $\xi^{(i)}$ by the variables $\xi^{(2)}, \xi^{(3)}, \ldots, \xi^{(n)}$.

For application in forecast problems, let us interpret (138) in terms of conditioned expectations. Writing $(C) = (\xi^{(1)} = \xi_1, \ldots, \xi^{(n)} = \xi_n)$, an estimate of $\xi_C^{(0)}$, which is the best linear one in the sense of the principle of the least squares, is yielded by

(149) $$F_C[\xi^{(0)}] = m_0 + a_{1n}(\xi_1 - m_1) + \cdots + a_{nn}(\xi_n - m_n).$$

We have

$$\eta_C^{(0)} = \xi_C^{(0)} - F_C[\xi^{(0)}]$$

and

$$E[\xi_C^{(0)} - F_C[\xi^{(0)}]]^2 = E[\eta_C^{(n)}]^2 = D^2(\eta^{(n)}),$$

where the latter relation results from (57).

If the variables $\xi^{(1)}, \ldots, \xi^{(n)}$ are mutually uncorrelated, the coefficients a_{ik} are independent of k. Writing $a_{ik} = a_i$, and taking $(C) = (\xi^{(i_1)} = \xi_{i_1} \ldots, \xi^{(i_k)} = \xi_{i_k})$, we get in this case

(150) $$F_C[\xi^{(0)}] = m_0 + a_{i_1}(\xi_{i_1} - m_{i_1}) + \cdots + a_{i_k}(\xi_{i_k} - m_{i_k}).$$

Further, if $\eta^{(n)}$ is independent of the variables $\xi^{(1)}, \ldots, \xi^{(n)}$ we have

(151) $$E_C[\eta^{(n)}] = E[\eta^{(n)}] = 0,$$

and

(152) $$E_C[\xi^{(0)}] = m_0 + a_{1n}(\xi_1 - m_1) + \cdots + a_{nn}(\xi_n - m_n) + E_C[\eta^{(n)}] = F_C[\xi^{(0)}].$$

19. **Linear autoregression analysis of the discrete stationary process.**

In this section, the linear regression analysis as surveyed in the previous section will be applied to the variables $\xi(t)$ connected with a stationary process $\{\xi(t)\}$. Approximating $\xi(t)$ by means of $\xi(t-1), \ldots, \xi(t-n)$, a well-defined procedure of consecutive approximations will be given. After a passage to the limit, we shall arrive at a residual variable with properties corresponding to the case of a finite number of approximations.

Let $\{\xi(t)\}$ be a stationary process with finite dispersion σ, with mean m, and with principal correlation determinants $\Delta(r, n)$ given by (81). Approximating $\xi(t)$ by means of $\xi(t-1, \ldots, t-n) = [\xi(t-1), \ldots, \xi(t-n)]$ as described in the previous section, let the residual variable be given by

$$(153)\qquad \eta(t;n) = \xi(t) - m - a(1,n)\cdot[\xi(t-1)-m] - a(2,n)\cdot \\ \cdot[\xi(t-2)-m] - \cdots - a(n,n)\cdot[\xi(t-n)-m].$$

In case $\Delta(r, n-1) \neq 0$, formula (139) yields

$$(154)\qquad D^2(\xi(t)) \geq D^2(\eta(t;n)) = \sigma^2\cdot\Delta(r,n)/\Delta(r,n-1) \geq 0.$$

Since $D^2(\eta(t;n))$, $n = 1, 2, \ldots$, forms a non-increasing sequence (cf. (147)), we get

$$(155)\qquad 1 \geq \frac{\Delta(r,1)}{1} \geq \frac{\Delta(r,2)}{\Delta(r,1)} \geq \cdots \geq \frac{\Delta(r,n)}{\Delta(r,n-1)} \geq \cdots \geq 0.$$

According to the analysis in section 14, one of two cases is present. Either $\Delta(r,n)$ is above zero for all n, or $\Delta(r,n)$ is > 0 for $n < h$ while $\Delta(r,n) = 0$ for $n \geq h$. The number h appearing in the latter case equals the rank of linear singularity of the process considered. Paying regard to these facts, and to the relation (155), it is evident that any stationary process belongs to one, and only to one, of the following classes:

(I). The process is non-singular, and there exists a positive constant $\varkappa^2 \leq 1$ such that

$$(156)\qquad D^2(\eta(t;n))/\sigma^2 = \Delta(r,n)/\Delta(r,n-1) \to \varkappa^2 \leq 1 \text{ as } n\to\infty.$$

(II). The process is singular, say of rank h.

(III). The process presents no singularity of finite rank, but

$$(157)\qquad D^2(\eta(t;n))/\sigma^2 = \Delta(r,n)/\Delta(r,n-1) \to 0 \text{ as } n\to\infty.$$

In this case, the process will be termed *singular of infinite rank.*[10]

In the following analysis, we shall assume that the process considered belongs to the first class.

Autoregression analysis of a non-singular process. According to (140), the variables (153) may be written on the form

$$(158)\qquad \eta(t;n) = \xi(t) - m - c_1\,\xi_1^*(t) - c_2\,\xi_2^*(t) - \cdots - c_n\,\xi_n^*(t),$$

where the standardized variables $\xi_k^*(t)$ are given by (cf. (142))

$$(159)\qquad \xi_n^*(t) = \frac{\begin{vmatrix} 1 & , r_1 & ,\ldots, & r_{n-2}, & \xi(t-1)-m \\ r_1 & , 1 & ,\ldots, & r_{n-3}, & \xi(t-2)-m \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{n-1}, & r_{n-2}, & \ldots, & 1 & , \xi(t-n)-m \end{vmatrix}}{\sigma\cdot\sqrt{\Delta(r,n-1)\cdot\Delta(r,n-2)}},$$

and where the coefficients c_i are uniquely determined, and independent of n (cf. (145)),

$$c_i = r\,[\xi(t),\, \xi_i^*(t)] \cdot \sigma. \tag{160}$$

According to the general analysis in section 13, the variables $\xi_k^*(t)$ will, for an arbitrarily fixed k, constitute a stationary process $\{\xi_k^*(t)\}$. It is seen that the variables $\xi_k^*(t)$ that form this process will be of type (61). Similarly, for any n the variables $\eta(t;n)$ will constitute a stationary process $\{\eta(t;n\}$ of the same type (61). It will next be shown that the sequence $\{\eta(t;n)\}$ is convergent in probability as $n \to \infty$.

According to (141) and (156) we have in the first place

$$D^2(\eta(t;n)) = \sigma^2 - c_1^2 - \cdots - c_n^2 \to \varkappa^2\sigma^2 \text{ as } n \to \infty.$$

We conclude that Σc_n^2 is convergent, and that

$$c_1^2 + c_2^2 + \cdots = (1 - \varkappa^2)\,\sigma^2. \tag{161}$$

Next, the variables $\xi_1^*(t), \xi_2^*(t), \ldots$ being uncorrelated (cf. (144)), we get

$$D^2(\eta(t;n+p) - \eta(t;n)) = c_{n+1}^2 + \cdots + c_{n+p}^2.$$

Keeping in mind that Σc_i^2 is convergent, it follows that this dispersion tends to zero uniformly in p as $n \to \infty$. Hence, paying regard to a remark in section 13 (see p. 40 f.), we conclude that the sequence $\eta(t;1), \eta(t;2), \ldots.$ is convergent in probability. According to theorem 1, this implies that the sequence

$$\{\eta(t;1)\},\ \{\eta(t;2)\}, \ldots. \tag{162}$$

will also converge in probability.

The limit process of the sequence (162) will be denoted $\{\eta(t)\}$. In analogy to the case of a finite number of approximations, the process $\{\eta(t)\}$ and the corresponding variables $\eta(t_1, \ldots, t_n)$ will be termed *residual*. According to a remark on p. 41, the mean and dispersion of the residual process will be given by the corresponding limit characteristics of the sequence (162).

Observing that $\eta(t;n)$ is non-correlated with the variables $\xi(t-1), \ldots, \xi(t-n)$, and keeping in mind that the dispersion of $\eta(t;n)$ lies above a positive constant, it follows readily that the limit residual $\eta(t)$ is non-correlated with $\xi(t-n)$ for any positive

integer n. Hence we conclude that $\eta(t)$ for any $k \geq 0$ is uncorrelated with all of the variables $\xi_1^*(t-k)$, $\xi_2^*(t-k), \ldots$ defined by (159) (cf. (137)). Further we have

$$E[(\xi(t)-m)\cdot\eta(t)] = E[(\xi(t)-m-c_1\xi_1^*(t)-\cdots-c_n\xi_n^*(t))\cdot\eta(t)] =$$

$$= \lim_{n\to\infty} E[\eta(t;n)\cdot\eta(t)] = E[\eta(t)\cdot\eta(t)] = D^2(\eta(t)).$$

Hence, paying regard to the relations (cf. p. 77)

$$E[\eta(t)] = E[\eta(t;n)] = 0,$$

we get

$$r(\xi(t),\eta(t)) = D(\eta(t))\,/\,D(\xi(t)) = \varkappa.$$

Writing generally

$$r(\xi(t+n),\eta(t))\,/\,\varkappa = b_n, \tag{163}$$

we thus have $b_0 = 1$, and $b_n = 0$ for $n < 0$.

The above-mentioned properties of the residuals correspond directly with the finite case dealt with in the previous section. There is also another important analogy. Considering in the finite case the residuals obtained when approximating $\xi^{(0)}$ and $\xi^{(p)}$ by $[\xi^{(1)},\ldots,\xi^{(n)}]$ and $[\xi^{(p+1)},\ldots,\xi^{(n)}]$ respectively, a short reflection shows that these residuals are non-correlated. In order to show, correspondingly, that the residuals $\eta(t)$ and $\eta(t-p)$ are non-correlated, let it be observed that

$$E[\eta(t)\cdot\eta(t-p)] = \lim_{k\to\infty} E\,[\eta(t;k)\cdot[\xi(t-p)-m-c_1\,\xi_1^*(t-p)-\cdots-$$
$$-\,c_{k-p+1}\,\xi_{k-p+1}^*(t-p)-\cdots-c_k\,\xi_k^*(t-p)]].$$

Denoting by ϱ_{k-p} the sum appearing in the second row, we have $E[\varrho_{k-p}] = 0$, and $D^2(\varrho_{k-p}) \to 0$ for any $p > 0$ as $k\to\infty$. Thus

$$r(\eta(t),\eta(t-p)) = \lim_{k\to\infty} r(\eta(t;k),\ \xi(t-p)-m-c_1\,\xi_1^*(t-p)-\cdots-$$
$$-\,c_{k-p}\,\xi_{k-p}^*(t-p)).$$

According to previous remarks, the correlation coefficient in the right member equals zero for any $p > 0$. Observing that

$$r(\eta(t),\eta(t+p)) = r(\eta(t-p),\eta(t)) = r(\eta(t),\eta(t-p)) = 0 \tag{164}$$

we conclude that the process $\{\eta(t)\}$ is non-autocorrelated.

Summing up the results, we get the following theorem.

Theorem 6. A residual process $\{\eta(t)\}$ obtained from a non-singular stationary process $\{\xi(t)\}$ is stationary and non-autocorrelated. The variable $\eta(t)$ is non-correlated with $\xi(t-1), \xi(t-2), \ldots,$ while

$$r(\xi(t), \eta(t)) = D(\eta(t)) / D(\xi(t)).$$

The arguments used in the proof of this theorem also apply in the remaining cases II and III. As the residual variables $\eta(t)$ are here seen to be vanishing, their correlation properties will be indeterminate. Accordingly, these cases need no further comment.

Illustrations of the autoregression analysis of stationary processes will be given in sections 25 and 26 (cf. also p. 92).

20. A canonical form of the discrete stationary process.

In this section it will be shown that the residual processes arrived at in the preceding section give a basis for the construction of a canonical form of the stationary process with finite dispersion.

Until further notice, the random variables considered will be assumed to have a vanishing mean.

In the case of a finite number of approximations dealt with in section 18, the following representation holds for the variable $\xi^{(0)}$ subjected to the regression analysis (see formula (138))

$$\xi^{(0)} = \eta^{(n)} + a_1 \xi^{(1)} + a_2 \xi^{(2)} + \cdots + a_n \xi^{(n)}. \tag{165}$$

Considering, on the other hand, the residuals $\eta(t; k)$ defined by (153) and obtained from a stationary process $\{\xi(t)\}$, it may be (cf. (148)) that the minimizing coefficients $a(i, k)$ are bounded in modulus by a constant not surpassing $1/\varkappa$. In such a case, a diagonal selection procedure will show that there exist a real sequence $a_1, a_2, \ldots$ and a sequence of integers $k_1, k_2, \ldots$ such that for all i

$$\lim_{s \to \infty} a(i, k_s) = a_i, \qquad |a_i| \leq 1/\varkappa.$$

Hence, we are led to ask if for these coefficients a_i

$$\lim_{n \to \infty} [\eta(t) + a_1 \xi(t-1) + \cdots + a_n \xi(t-n)]$$

exists and equals $\xi(t)$, a relation which would correspond to (165). If the answer is affirmative, a non-singular stationary process $\{\xi(t)\}$ could always be written on the form

$$\{\eta(t)\} + a_1\{\xi(t-1)\} + a_2\{\xi(t-2)\} + \ldots, \tag{166}$$

where $\{\eta(t)\}$ is non-autocorrelated, and $\eta(t)$ is uncorrelated with $\xi(t-1)$, $\xi(t-2)$, etc.

However, as will be seen in section 26, certain conditions would have to be imposed upon $\{\xi(t)\}$ in order to secure the representation (166). For the present we shall leave open all questions in this matter, and proceed to an aspect of the finite case suggesting another canonical form of the stationary process.

Writing

$$\vartheta(t;k) = \xi(t) - \eta(t;k),$$

it is evident that the sequence $\{\vartheta(t;k)\}$ is convergent as $k \to \infty$. Denoting the limit process by $\{\vartheta(t)\}$, we obtain

$$\{\xi(t)\} = \{\eta(t)\} + \{\vartheta(t)\}.$$

Since $\vartheta(t;k)$ is a linear expression in $\xi(t-1), \ldots, \xi(t-k)$, and thus uncorrelated with $\eta(t+n)$ for all $n \geq 0$, it follows that $\eta(t)$ is uncorrelated with the variables $\vartheta(t)$, $\vartheta(t-1)$, etc. So far there is a complete analogy with the finite relation (165). In further analogy, $\vartheta(t;k)$ can be written as a sum involving the uncorrelated residuals $\eta(t-1;k-1), \ldots, \eta(t-k+1;1)$, and $\eta(t-k;0) = \xi(t-k)$. This circumstance suggests the question of whether $\{\vartheta(t)\}$ is a linear expression in $\{\eta(t-1)\}$, $\{\eta(t-2)\}$, etc. Were the answer in the affirmative, then $\{\xi(t)\}$ could always be written on the form $\{\eta(t)\} + b_1\{\eta(t-1)\} + b_2\{\eta(t-2)\} + \ldots$, where $\{\eta(t)\}$ is non-autocorrelated. It will be found that such a sum will not be sufficient as a canonical form for the stationary process — in general, a singular process $\{\psi(t)\}$, which is uncorrelated with $\{\eta(t)\}$, has to be added in order to obtain $\{\xi(t)\}$.

After these introductory remarks, let $\{\xi(t)\}$ represent an arbitrary non-singular stationary process with finite dispersion σ, and with zero for mean value. In the first place, let an approximation procedure be performed on $\xi(t)$ by means of the residuals $\eta(t;k), \ldots, \eta(t-n;k)$ given by (158). Denoting the new residuals by $\psi(t;n;k)$ we write

(167) $\psi(t;n;k)=\xi(t)-b(0;n;k)\cdot\eta(t;k)-\cdots-b(n;n;k)\cdot\eta(t-n;k).$

Since $\xi(t)$ is non-singular, the coefficients b minimizing $D(\psi(t;n;k))$ will be uniquely determined.

The processes $\{\psi(t;n;k)\}$ defined by (167) are, like $\{\eta(t;k)\}$, of type (61). The processes $\{\psi(t;n;k)\}$ will next be subjected to a repeated passage to the limit, first in respect of k, and then in respect of n.

Letting k tend to infinity, and paying regard to the relation

$$\lim_{k\to\infty} r(\eta(t-p;k),\eta(t-q;k))=0,$$

which according to theorem 6 is valid for $p \neq q$, we get (cf. (163))

$$\lim_{k\to\infty} b(p;n;k)=r(\xi(t),\eta(t-p))/\varkappa=b_p, \tag{168}$$

independently of n. Thus, keeping in mind that the variable $\eta(t,t-1,..,t-n;k)$ tends to $\eta(t,t-1,..,t-n)$ as $k\to\infty$, it follows that for all n and t

$$\lim_{k\to\infty}\psi(t;n;k)=\xi(t)-\eta(t)-b_1\eta(t-1)-\cdots-b_n\eta(t-n). \tag{169}$$

Let the limit variables thus obtained be denoted by $\psi(t;n)$. Now, holding n fixed, the variables $\psi(t;n)$ will obviously constitute a stationary process $\{\psi(t;n)\}$ which is the limit of the sequence $\{\psi(t;n;1)\}, \{\psi(t;n;2)\},\ldots$

Keeping in mind that the variables $\eta(t)$ are mutually uncorrelated, a short calculation shows that

$$D^2(\psi(t;n))=[1-(1+b_1^2+\cdots+b_n^2)\cdot\varkappa^2]\,\sigma^2. \tag{170}$$

Concluding that the series Σb_i^2 is convergent, let us write

$$K^2=1+b_1^2+b_2^2+\ldots., \tag{171}$$

and further

$$\zeta(t;n)=\eta(t)+b_1\cdot\eta(t-1)+\cdots+b_n\cdot\eta(t-n). \tag{172}$$

Thus prepared, let n tend to infinity in (169). Since Σb_i^2 converges, we have

$$D^2(\zeta(t;n+p)-\zeta(t;n))=(b_{n+1}^2+\cdots+b_{n+p}^2)\cdot D^2(\eta)\to 0$$

uniformly in p as $n \to \infty$. It follows that the sum $\eta(t) + b_1 \eta(t-1) + b_2 \eta(t-2) + \cdots$ is convergent. Denoting the sum variable by $\zeta(t)$, we have

$$(173) \quad \zeta(t) = \lim_{n\to\infty} \zeta(t;n) = \eta(t) + b_1 \eta(t-1) + b_2 \eta(t-2) + \cdots$$

Further, the variables $\zeta(t)$ and $\psi(t) = \lim \psi(t;n)$ constitute two stationary processes $\{\zeta(t)\} = \lim_{n\to\infty} \{\zeta(t;n)\}$, and $\{\psi(t)\} = \lim_{n\to\infty} \{\psi(t;n)\}$ respectively.

Observing that (173) yields

$$(174) \qquad D^2(\zeta(t)) = (1 + b_1^2 + b_2^2 + \cdots) \cdot \varkappa^2 \sigma^2 = \varkappa^2 \cdot K^2 \cdot \sigma^2,$$

two cases may be distinguished:

(A) $\varkappa \cdot K = 1$. Then $D(\psi(t)) = 0$, and

$$(175) \qquad \{\xi(t)\} = \{\zeta(t)\};$$

(B) $\varkappa \cdot K < 1$. Then $D(\psi(t)) > 0$, and

$$(176) \qquad \{\xi(t)\} = \{\zeta(t)\} + \{\psi(t)\}.$$

Advancing that $\{\psi(t)\}$ is singular, it is seen that (176) covers both (175) and the cases II and III (see p. 81). Moreover, giving $\psi(t)$ the same mean as $\xi(t)$, the representation (176) evidently holds also in case $\{\xi(t)\}$ has a non-vanishing mean.

Formula (176) is the desired canonical form for a stationary process with finite dispersion. As already pointed out, the variable $\zeta(t)$ corresponds directly with the case of a variable ξ in a finite number of dimensions. Further, according to (164) and (173), our $\{\zeta(t)\}$ presents a certain similarity to the general process $\{\gamma(t)\}$ of linear regression as introduced in section 15 γ. However, $\{\zeta(t)\}$ is still more general, for the variables $\eta(t)$ constituting $\zeta(t)$ are non-correlated, while those forming $\gamma(t)$ moreover are independent.

Some characteristic properties of the variables $\zeta(t)$ and $\psi(t)$ appearing in the canonical formula (176) will now be proved, and the main results then comprehended in a theorem.

Observing that

$$\left(\sum_{i=0}^{\infty} b_{p+i} \cdot b_i\right)^2 \leq \left(\sum_{i=0}^{\infty} b_{p+i}^2\right) \cdot \left(\sum_{i=0}^{\infty} b_i^2\right) = K^2 \cdot \sum_{p}^{\infty} b_i^2 \to 0 \text{ as } p \to \infty,$$

we conclude that $\sum_{i=0}^{\infty} b_i \cdot b_{i+p}$ is convergent for all p, and that

$$(177) \qquad r_p(\zeta) = E[\zeta(t) \cdot \zeta(t-p)] / \varkappa^2 \cdot K^2 \cdot \sigma^2 =$$
$$= \lim_{n \to \infty} (b_p + b_{p+1} \cdot b_1 + \cdots + b_n \cdot b_{n-p}) / K^2 \to 0 \text{ as } p \to \infty.$$

Paying regard to (163), we get further

$$(178) \qquad r_p(\zeta) = (b_p + b_{p+1} \cdot b_1 + b_{p+2} \cdot b_2 + \cdots) / K^2 =$$
$$= \frac{1}{\varkappa K} \cdot \lim_{n \to \infty} r(\xi(t), \eta(t-p) + b_1 \cdot \eta(t-p-1) + \cdots + b_n \cdot \eta(t-p-n)) =$$
$$= r(\xi(t), \zeta(t-p)) / \varkappa \cdot K.$$

Considering in the second place $\{\psi(t)\}$, we obtain from (170)

$$(179) \qquad D^2(\psi(t)) = (1 - \varkappa^2 \cdot K^2) \cdot \sigma^2,$$

Next, in case $D(\psi) > 0$ we get from (169)

$$r(\psi(t; n); \eta(t+p)) = r(\xi(t) - \eta(t) - b_1 \cdot \eta(t-1) - \cdots - b_n \cdot \eta(t-n); \eta(t+p)).$$

Keeping in mind that $\eta(t)$ is uncorrelated with any $\xi(t-p)$ and with any $\eta(t \pm p)$ for $p \neq 0$, and paying regard to (163), a short calculation shows that $r(\psi(t; n), \eta(t+p)) = 0$ for $p \geq -n$. It follows that for any $p \gtreqless 0$

$$r(\psi(t), \eta(t+p)) = \lim_{n \to \infty} r(\psi(t; n), \eta(t+p)) = 0.$$

Hence the fundamental relation

$$(180) \qquad r(\psi(t), \zeta(t+p)) = \lim r(\psi(t), \zeta(t+p; n)) = 0, \qquad p \gtreqless 0,$$

which shows that the processes $\{\psi(t)\}$ and $\{\zeta(t)\}$ are non-correlated. Thus we have (cf. (59) and (119))

$$(181) \qquad r_k(\xi) = \frac{D^2(\psi)}{D^2(\xi)} \cdot r_k(\psi) + \frac{D^2(\zeta)}{D^2(\xi)} \cdot r_k(\zeta).$$

For the preparation of the remaining proof of the singularity of $\{\psi(t)\}$, let it be observed that the non-correlation between $\eta(t)$ and $\eta(t+p)$ for $p \neq 0$ implies that for any real number a_1

$$D^2(\zeta(t; n) + a_1 \cdot \zeta(t-1; n)) = D^2(\eta(t) + (a_1 + b_1) \cdot \eta(t-1) +$$
$$+ (a_1 \cdot b_1 + b_2) \cdot \eta(t-2) + (a_1 \cdot b_2 + b_3) \cdot \eta(t-3) + \cdots + a_1 \cdot b_n \cdot \eta(t-n-1)) =$$
$$= D^2(\eta(t) + B_1 \cdot \eta(t-1) + \cdots + B_{n+1} \cdot \eta(t-n-1)) =$$
$$= (1 + B_1^2 + B_2^2 + \cdots + B_{n+1}^2)\,\varkappa^2 \cdot \sigma^2 \geq \varkappa^2 \cdot \sigma^2,$$

the auxiliary constants B_m introduced being real. By the same argument we conclude that for any real $a_1, a_2, \ldots, a_p$

$$(182) \qquad D^2(\zeta(t) + a_1 \cdot \zeta(t-1) + \cdots + a_p \cdot \zeta(t-p)) =$$
$$= \lim_{j \to \infty} D^2(\zeta(t; n_j) + a_1 \cdot \zeta(t-1; n_j) + \cdots + a_p \cdot \zeta(t-p; n_j)) \geq \varkappa^2 \cdot \sigma^2.$$

Considering now the variable $\xi(t) + a_1 \cdot \xi(t-1) + \cdots + a_p \cdot \xi(t-p)$, and paying regard to (176) and (180), a short calculation shows that

$$(183) \qquad D^2(\xi(t) + a_1 \cdot \xi(t-1) + \cdots + a_p \cdot \xi(t-p)) =$$
$$= D^2(\psi(t) + a_1 \cdot \psi(t-1) + \cdots + a_p \cdot \psi(t-p)) +$$
$$+ D^2(\zeta(t) + a_1 \cdot \zeta(t-1) + \cdots + a_p \cdot \zeta(t-p)).$$

However, if an $\varepsilon > 0$ is arbitrarily given, (156) implies that a number $p(\varepsilon)$ and a real sequence, say $a_1 = a_1^*, a_2 = a_2^*, \ldots, a_p = a_p^*$, exist such that the left member of (183) is less than $\varkappa^2 \cdot \sigma^2 + \varepsilon$. On the other hand, (182) shows that the second variance in the right member of (183) is not below $\varkappa^2 \cdot \sigma^2$. Hence it follows that

$$D^2(\psi(t) + a_1^* \cdot \psi(t-1) + \cdots + a_p^* \cdot \psi(t-p)) \leq \varepsilon.$$

Since ε is arbitrary, this relation implies that $\{\psi(t)\}$ is singular of a finite or infinite rank.

Summing up, we have the following theorem in which one of the variables $\{\psi(t)\}$ and $\{\zeta(t)\}$ may be vanishing.[11]

Theorem 7. Denoting by $\{\xi(t)\}$ an arbitrary discrete stationary process with finite dispersion, there exists a three-dimensional stationary process $\{\psi(t), \zeta(t), \eta(t)\}$ with the following properties:

(*A*) $\{\xi(t)\} = \{\psi(t)\} + \{\zeta(t)\}$.
(*B*) *$\{\psi(t)\}$ and $\{\zeta(t)\}$ are non-correlated.*
(*C*) *$\{\psi(t)\}$ is singular.*
(*D*) *$\{\eta(t)\}$ is non-autocorrelated, and $E[\eta(t)] = E[\zeta(t)] = 0$.*
(*E*) $\{\zeta(t)\} = \{\eta(t)\} + b_1 \cdot \{\eta(t-1)\} + b_2 \cdot \{\eta(t-2)\} + \cdots$
where b_n represent real numbers such that Σb_n^2 is convergent.

Illustrations. In order to illustrate the autoregression analysis, let us consider a normal stationary process $\{\xi(t)\}$ as defined by the characteristic function (111). Assuming for the sake of formal simplicity that $m = 0$, and that $\sigma = 1$, we shall first investigate a sum variable $\zeta[t]$ of type (61).

Writing $\zeta(t, t-1, \ldots, t-n) = [\zeta(t), \zeta(t-1), \ldots, \zeta(t-n)]$, we have by definition

$$\begin{cases} \zeta(t) = \xi(t) + a_1\,\xi(t-1) + \cdots + a_h\,\xi(t-h) \\ \zeta(t-1) = \xi(t-1) + a_1\,\xi(t-2) + \cdots + a_h\,\xi(t-h-1) \\ \cdots\cdots\cdots \\ \zeta(t-n) = \xi(t-n) + a_1\,\xi(t-n-1) + \cdots + a_h\,\xi(t-n-h). \end{cases}$$

According to the introductory remarks in section 14, the characteristic function of $\zeta(t, t-1, .., t-n)$, say $f_n^*(Z_t, Z_{t-1}, .., Z_{t-n}) = e^{-\frac{1}{2}Q_n^*}$, will for all $n \lesseqgtr h$ be obtained from the characteristic function $f_{n+h}(X_t, X_{t-1}, .., X_{t-n-h})$ by the substitution

$$(184)\quad \begin{cases} X_t = Z_t \\ X_{t-1} = a_1 Z_t + Z_{t-1} \\ \cdots\cdots \\ X_{t-h} = a_h Z_t + a_{h-1} Z_{t-1} + \cdots + Z_{t-h} \\ X_{t-h-1} = a_h Z_{t-1} + \cdots + a_1 Z_{t-h} + Z_{t-h-1} \\ \cdots\cdots\cdots \\ X_{t-n} = a_h Z_{t-n+h} + \cdots + a_1 Z_{t-n+1} + Z_{t-n} \\ X_{t-n-1} = a_h Z_{t-n+h-1} + \cdots + a_2 Z_{t-n+1} + a_1 Z_{t-n} \\ \cdots\cdots\cdots \\ X_{t-n-h+1} = a_h Z_{t-n+1} + a_{h-1} Z_{t-n} \\ X_{t-n-h} = a_h Z_{t-n} \end{cases}$$

We conclude that the distribution of $\zeta(t, t-1, .., t-n)$ is normal, and that we obtain Q_n^* from Q_{n+h} by the substitution (184).

According to the substitution theory of quadratic forms, the matrix defining Q_n^* is obtained as follows (see e.g. G. Kowalewski (1909), § 94). Writing $A_n(\xi)$ for the matrix of Q_n, and $B_{h,n}$ for the matrix of the substitution (184), we have

$$A_{n+h}(\xi) = \begin{pmatrix} 1, & r_1, & r_2, & \ldots, & r_{n+h} \\ r_1, & 1, & r_1, & \ldots, & r_{n+h-1} \\ r_2, & r_1, & 1, & \ldots, & r_{n+h-2} \\ \cdots & & & & \\ r_{n+h}, & r_{n+h-1}, & r_{n+h-2}, & \ldots, & 1 \end{pmatrix},$$

$$B_{h,n} = \begin{pmatrix} 1, & 0, & 0, & \ldots, & 0, & 0, \ldots, & 0 \\ a_1, & 1, & 0, & \ldots, & 0, & 0, \ldots, & 0 \\ \cdots & & & & & & \\ a_h, & a_{h-1}, & a_{h-2}, & \ldots, & 1, & 0, \ldots, & 0 \\ 0, & a_h, & a_{h-1}, & \ldots, & a_1, & 1, \ldots, & 0 \\ \cdots & & & & & & \\ 0, & 0, & \ldots\ldots, & a_h, & a_{h-1}, \ldots, & a_1, & 1 \\ \cdots & & & & & & \\ 0, & 0, & \ldots\ldots\ldots, & & 0, & a_h, & a_{h-1} \\ 0, & 0, & \ldots\ldots\ldots, & & 0, & 0, & a_h \end{pmatrix}$$

In the first place we must form a product matrix consisting of $n+h+1$ rows and $n+1$ columns, and take for the i^{th} element in the j^{th} row the inner product of the j^{th} row in $A_{n+h}(\xi)$ by the i^{th} column in $B_{n,h}$. Multiplying then $B_{h,n}$ and the product matrix by columns, we arrive at the matrix required. Denoting this by $A_n(\zeta)$, and by M' the transposed of a matrix M, we have

$$A_n(\zeta) = B'_{h,n} \cdot A_{n+h}(\xi) \cdot B_{h,n}.$$

Forming in the same way the infinite matrix

(185) $$A(\zeta) = B' \cdot A(\xi) \cdot B,$$

where

(186) $$A(\xi) = \begin{pmatrix} 1, & r_1, & r_2, & \dots \\ r_1, & 1, & r_1, & \dots \\ r_2, & r_1, & 1, & \dots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \quad B = \begin{pmatrix} 1, & 0, & 0, & 0, & \dots & & & & \\ a_1, & 1, & 0, & 0, & \dots & & & & \\ a_2, & a_1, & 1, & 0, & \dots & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & & & \\ a_h, & a_{h-1}, & a_{h-2}, & \dots, & a_1, & 1, & 0, & 0, & \dots \\ 0, & a_h, & a_{h-1}, & a_{h-2}, & \dots, & a_1, & 1, & 0, & \dots \\ 0, & 0, & a_h, & a_{h-1}, & \dots, & \dots, & a_1, & 1, & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix},$$

it is readily verified that $A_n(\zeta)$ equals the principal minor of order $n+1$ in $A(\zeta)$. We conclude that the distribution of $\zeta(t, t-1, \dots)$ is normal, and that $A(\zeta)$ is the matrix of the infinite quadratic form $Q^*(Z_t, Z_{t-1}, \dots)$ appearing in the exponent of the characteristic function of $\zeta(t, t-1, \dots)$.

Formally, the above procedure applies even in the case of an infinite sequence $(a_1, a_2, \dots)$. It is seen that if the double series appearing in the matrix Q^* are absolutely convergent, the variable $\zeta(t) = \sum_{i=0}^{\infty} a_i \xi(t-i)$ will be well-defined, and constitute a normal process. The characteristic function of the variable $\zeta(t, t-1, \dots)$ will be given by $e^{-\frac{1}{2}Q^*}$.

Next we shall consider a few particular instances.

Let $(1, b_1, b_2, \dots)$ represent a real sequence such that $K^2 = 1 + \sum b_i^2$ is finite, and let $\{\eta(t)\}$ be a normal and purely random process with vanishing mean, and dispersion equalling unity. Considering the variable $\zeta(t) = \eta(t) + b_1\eta(t-1) + b_2\eta(t-2) + \cdots$, it is readily verified that the above substitution procedure gives $A(\zeta) = B' \cdot A(\eta) \cdot B$, where

$$A(\eta) = \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot \\ 0 & 1 & 0 & \cdot & \cdot \\ 0 & 0 & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \quad B = \begin{pmatrix} 1, & 0, & 0, & 0, & 0, & \dots \\ b_1, & 1, & 0, & 0, & 0, & \dots \\ b_2, & b_1, & 1, & 0, & 0, & \dots \\ b_3, & b_2, & b_1, & 1, & 0, & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \quad A(\zeta) = \begin{pmatrix} K^2 & K^2 r_1, & K^2 r_2, & \dots \\ K^2 r_1, & K^2 & K^2 r_1, & \dots \\ K^2 r_2, & K^2 r_1, & K^2 & \dots \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix},$$

having in the latter matrix written r_k for $(b_k + b_1 b_{k+1} + b_2 b_{k+2} + \cdots)/K^2$. These results are seen to be in full agreement with formula (177), and with the correlation properties of a normal stationary process (see p. 62).

Considering in the second place a process $\{\xi(t)\}$ which is singular of rank h, there exists, by definition, a sequence $(a_1, \ldots, a_h)$ such that the relation (77) is satisfied. Accordingly, the infinite matrix $A(\zeta) = B' \cdot A(\xi) \cdot B$ formed by means of the matrices (186), will consist entirely of zeros. For instance, letting $\xi(t) + \xi(t-1) = 0$ be the relation of singularity, we have $h = a_1 = 1$ (cf. (107)). A short calculation will show that

$$A(\xi) = \begin{pmatrix} 1, & -1, & 1, & -1, & \ldots \\ -1, & 1, & -1, & 1, & \ldots \\ 1, & -1, & 1, & -1, & \ldots \\ -1, & 1, & -1, & 1, & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots \\ 1 & 1 & 0 & 0 & \ldots \\ 0 & 1 & 1 & 0 & \ldots \\ 0 & 0 & 1 & 1 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \quad A(\zeta) = \begin{pmatrix} 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

If $\{\xi(t)\}$ is singular of infinite rank, there exists, for every integer n and every $\varepsilon > 0$, a number $h(\varepsilon, n)$ and a sequence $a_i(\varepsilon, n)$ such that the variable $\xi(t) + a_1 \xi(t-1) + \cdots + a_h \xi(t-h)$ will give rise to a matrix Q_n^*, whose elements are all less than ε in modulus.

Proceeding to the operation of summing independent processes, let $\{\zeta(t)\}$ and $\{\psi(t)\}$ stand for two independent normal processes. Denoting the sum process by $\{\xi(t)\}$, and indicating all symbols referring to the three processes by ξ, ζ and ψ respectively, and paying regard to the evident fact that the characteristic function of $\xi(t, t-1, \ldots)$ is the mathematical product of the characteristic functions of the variables $\zeta(t, t-1, \ldots)$ and $\psi(t, t-1, \ldots)$, we obtain $D^2(\xi) \cdot Q(\xi) = D^2(\zeta) \cdot Q(\zeta) + D^2(\psi) \cdot Q(\psi)$, i.e.

$$D^2(\xi) \cdot \sum_0^\infty \sum_0^\infty r_{|p-q|}(\xi) \cdot X_{t-p} X_{t-q} = D^2(\zeta) \cdot \sum_0^\infty \sum_0^\infty r_{|p-q|}(\zeta) \cdot X_{t-p} X_{t-q} + D^2(\psi) \cdot \sum_0^\infty \sum_0^\infty r_{|p-q|}(\psi) \cdot X_{t-p} X_{t-q}.$$

In full agreement with (59) we obtain the relation (181), and observe that mutually uncorrelated normal processes are always independent.

The simple types of normal process mentioned above are sufficient to illustrate theorem 7. Starting from a purely random normal process $\{\eta(t)\}$ with suitable dispersion, forming a sum process $\{\zeta(t)\}$ of type $\{\eta(t)\} + b_1 \{\eta(t-1)\} + b_2 \{\eta(t-2)\} + \ldots$, and adding an independent normal process $\{\psi(t)\}$ ruled by an appropriate singularity, we shall arrive at an arbitrarily prescribed, normal and stationary process $\{\xi(t)\}$.

CHAPTER III.

On the theory of some special stationary processes.

21. On the concept of stochastical difference equation.

The relation

$$(187) \qquad \{\xi(t)\} + a_1 \cdot \{\xi(t-1)\} + \cdots + a_h \cdot \{\xi(t-h)\} = \{\eta(t)\}$$

arrived at in section 15 δ presents a formal analogy with an ordinary, or functional, difference equation

$$(188) \qquad x(t) + a_1 \cdot x(t-1) + \cdots + a_h \cdot x(t-h) = y(t).$$

We have found under certain conditions concerning the coefficients a_i, and in case the process $\{\eta(t)\}$ is purely random, that there exists a stationary process $\{\xi(t)\}$ which satisfies (187) and is of type

$$(189) \qquad \{\xi(t)\} = \{\eta(t)\} + b_1 \cdot \{\eta(t-1)\} + b_2 \cdot \{\eta(t-2)\} + \ldots$$

On the other hand, since under general conditions a solution of (188) will be of the form

$$(190) \qquad x(t) = y(t) + b_1 \cdot y(t-1) + b_2 \cdot y(t-2) + \ldots,$$

a clear formal analogy may be seen between (187) and (188) also in respect of the solutions.

Expressing the situation in words, a solution of type (190) of a functional difference equation is a moving average performed on the function $y(t)$ in the right member. Correspondingly, any sample series $(\xi_t, \xi_{t-1}, \ldots)$ connected with a process $\{\xi(t)\}$ of type (189) may be looked upon as a moving average of a purely random series $(\eta_t, \eta_{t-1}, \ldots)$.

Because of this parallelism, I propose to call (187) a *stochastical difference relation* between the processes $\{\xi(t)\}$ and $\{\eta(t)\}$. If $\{\eta(t)\}$ is known, and $\{\xi(t)\}$ unknown, (187) will be termed a *stochastical*

difference equation. An interpretation in the language of the theory of oscillatory mechanisms will reveal some interesting connexions between functional and stochastical difference equations, and exemplify the wide applicability of the new concept.

An oscillatory mechanism presents certain intrinsic features relevant to the structure of the movement considered. Studying the movement in integral time points, these features are summed up in the relation

$$(191) \qquad x(t) + a_1 \cdot x(t-1) + \cdots + a_h \cdot x(t-h) = 0.$$

Interpreting (191) as an ordinary difference equation, the solutions (see section 6) describe how the phenomenon would develop out from any initial values, say $x(t-1) = x_{t-1}, \ldots, x(t-h) = x_{t-h}$, if there were no external influence present.

In the ordinary difference equations, the external factors are taken into account by means of the function $y(t)$. Thus, instead of the value $-a_1 \cdot x_{t-1} - a_2 \cdot x_{t-2} - \cdots - a_h \cdot x_{t-h}$ to be expected for $x(t)$ when the earlier values are known, the variable in question takes on the value $y(t) - a_1 \cdot x_{t-1} - \cdots - a_h \cdot x_{t-h}$. In this approach the external influence is dealt with as functional, i. e. uniquely determined at any future time point.

The stochastical approach differs in the allowance for the external influence upon the mechanism. Here the external factors are not dealt with as functionally determined; they are only assumed to be ruled by certain probability laws. These laws constitute the stochastical process $\{\eta(t)\}$; in general, the probability laws are subjected only to the conditions (53)—(54) which express that the laws must not contradict themselves (cf. p. 3). The simple case investigated in section 15 δ corresponds to a purely random effect of the external factors, but nothing prevents us from approximating the external influence by a non-purely-random, or even a non-stationary process $\{\eta(t)\}$.

Having fixed a $\{\eta(t)\}$, any sample series $(\ldots, \eta_{t-1}, \eta_t, \eta_{t+1}, \ldots)$ will describe an actual realization of the external development. Since the movement of the mechanism is known when the external factors are determined, the sample series $(\ldots, \eta_{t-1}, \eta_t, \eta_{t+1}, \ldots)$ considered will correspond to a certain sample series, say $(\ldots, \xi_{t-1}, \xi_t, \xi_{t+1}, \ldots)$, of the process $\{\xi(t)\}$. However, as we possess only probability knowledge of the actual path $(\ldots, \eta_{t-1}, \eta_t, \eta_{t+1}, \ldots)$ of the external factors, we can reach only probability laws about the

behaviour of the oscillatory mechanism. These probability laws concerning $(\ldots, \xi_{t-1}, \xi_t, \xi_{t+1}, \ldots)$ constitute the process $\{\xi(t)\}$, and form a solution of the stochastical difference equation.

By means of the probability laws found for the mechanism, it will be possible to give information as to the average behaviour of the phenomenon considered, i. e. as to the expectations referring to $\{\xi(t)\}$. It should be observed that we cannot say in advance that the conclusions as to the average behaviour will be identical to those drawn from the functional difference equation (191). Nevertheless, it has often been argued — more or less explicitly — that any intrinsic tendency of the mechanism to produce periodic oscillations will, on the average, give rise to a corresponding oscillation in the phenomenon when influenced by random shocks (see e. g. SIR G. WALKER (1931), p. 522, and R. FRISCH (1933), p. 202). However, the following analysis will show that there are important instances when such inference based on analogy is incorrect, qualitatively as well as quantitatively. For instance, let $C \cdot q^t \cdot \cos(\lambda_1 \cdot t + \varphi)$ represent a solution of (191), i. e. a damped oscillation characteristic of the mechanism, and consider the simple case of purely random external shocks. Then, even if there is no other intrinsic tendency to oscillation present, a periodogram analysis for the search of the frequency λ_1 will be more or less misleading. As will be shown in section 25, there is in general a systematic deviation between λ_1 and the abscissa for which the expectation of the periodogram ordinate presents a maximum. It may even happen that there is no maximum at all in the neighbourhood of λ_1.

As soon as external influence cannot be considered free from random elements, the stochastical difference equation should be preferred to the functional equation (cf. the quotations from G. U. YULE (1927) in section 10). It is also obvious that the former embraces the latter as a special case, for any function $y(t)$ may be interpreted as a singular random process. Thus, the stochastical difference equations seem to merit particular interest.

When omitting all dispensable conditions as to $\{\eta(t)\}$ and the a_i's, the solutions of the stochastical difference equations become of a very general type, and embrace fundamentally different classes of random process. Having already seen in section 15 δ that the solutions cover the stationary processes of linear autoregression, let us in the second place consider the special equation

$$(192) \qquad \{\xi(t)\} - \{\xi(t-1)\} = \{\eta(t)\},$$

taking as before a purely random process for $\{\eta(t)\}$. The solutions of this equation are seen to form a type case of the *discrete homogeneous process* (see e. g. H. CRAMÉR (1937), Ch. VIII). In sharp contrast to the stationary processes, the oscillations here tend to increase in amplitude as time goes on. We may express this fact by saying that the homogeneous process is *evolutive* (cf. p. 1). In the particular case (192), the process $\{\xi(t)\}$ cannot be assumed to have been in movement during an infinite past. Accordingly, and in contradistinction to the stationary case, the analysis of this equation generally has to be restricted to an interval of type $(t_0 \leq t \leq \infty)$.

As already pointed out, there are many problems calling for investigation in connexion with the general stochastical difference equation. The coming section is reserved for some groundwork concerning such equations. In accordance with the program of the present study, non-stationary solutions will be dealt with only very briefly.

22. Some fundamentals concerning stochastical difference equations.

According to the definition given in the previous section,

$$(193) \qquad \{\xi(t)\} + a_1 \cdot \{\xi(t-1)\} + \cdots + a_h \cdot \{\xi(t-h)\} = \{\eta(t)\}$$

forms a stochastical difference equation in $\{\xi(t)\}$ if the coefficients a_i are real, and if $\{\eta(t)\}$ is a discrete random process. If $a_h \neq 0$, the equation will be termed of *order* h.

Let first an equation of order h with vanishing right member be considered,

$$(194) \qquad \{\xi(t)\} + a_1 \cdot \{\xi(t-1)\} + \cdots + a_h \cdot \{\xi(t-h)\} = 0.$$

If there are any solutions to this equation, these will be singular in the sense indicated in section 14, and have (77) with $m = 0$ for relation of singularity. Being particularly interested in stationary solutions with finite dispersion, it follows from the analysis in section 14 that there exists a non-vanishing stationary process which satisfies (194) if, and only if, the characteristic equation (34) has at least one root on the circumference of the unit circle.

It in seen that if $\{\xi(t)\}$ and $\{\psi(t)\}$ are stochastic processes such that $\{\xi(t)\}$ is a solution of (193), while $\{\psi(t)\}$ satisfies (194),

then $\{\xi(t)\}+\{\psi(t)\}$ will satisfy (193). This property of the stochastical difference equation forms another analogy to the functional case.

Secondly, we shall touch upon the case when the variable $\{\eta(t)\}$ appearing in (193) is of the type $\{\eta(t_0;\ t)\}=[\ldots, 0,\ 0,\ \eta(t_0),\ \eta(t_0+1),$ $\eta(t_0+2), \ldots]$ considered in connexion with theorem 1. It is evident that this equation has one, and only one, solution of type $\{\xi(t_0;\ t)\}=$ $=[\ldots, 0,\ 0,\ \xi(t_0),\ \xi(t_0+1), \ldots]$, and that this solution is given by the following system,

$$\begin{cases} \xi(t_0) = \eta(t_0), \\ \xi(t_0+1)=\eta(t_0+1)-a_1\cdot\xi(t_0)=\eta(t_0+1)-a_1\cdot\eta(t_0), \\ \xi(t_0+2)=\eta(t_0+2)-a_1\cdot\xi(t_0+1)-a_2\cdot\xi(t_0)=\eta(t_0+2)-a_1\cdot\eta(t_0+1)+ \\ \qquad\qquad +(a_1^2-a_2)\cdot\eta(t_0), \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

The general variable $\xi(t_0+t)$ is evidently of type

$$(195)\quad \xi(t_0+t)=\eta(t_0+t)+b_1\cdot\eta(t_0+t-1)+b_2\cdot\eta(t_0+t-2)+\cdots+b_t\cdot\eta(t_0).$$

The coefficients b_i introduced are seen to be identical to those used in section 15, δ. Thus, $b_1,\ b_2, \ldots, b_h$ will be obtained from the system (97), and the following ones from the difference equation (96). The following elementary theorem concerning the coefficients b_i will prove useful.[12]

Theorem 8. The series b_t defined by (96) *and* (97) *does not satisfy any difference equation of type* (32) *of lower order than h.*

Writing b_t on the form (33), let us examine the values, say $b(t)$, of this analytical function taken on for $t\leq 0$. Since every linear difference relation which is satisfied by this function for $t>0$ must hold also for $t\leq 0$ and *vice versa*, it is sufficient to verify theorem 8 for $t\leq 0$. According to the difference equation (96), we get $m=0$ and

$$(196)\begin{cases} b_1+a_1\cdot b(0)+a_2\cdot b(-1)+\cdots+a_{h-1}\cdot b(-h+2)+a_h\cdot b(-h+1)=0 \\ b_2+a_1\cdot b_1\ +a_2\cdot b(0)\quad +\cdots+a_{h-1}\cdot b(-h+3)+a_h\cdot b(-h+2)=0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ b_{h-1}+a_1\cdot b_{h-2}+a_2\cdot b_{h-3}+\cdots+a_{h-1}\cdot b(0)+a_h\cdot b(-1)=0 \\ b_h\quad +a_1\cdot b_{h-1}+a_2\cdot b_{h-2}+\cdots+a_{h-1}\cdot b_1\quad +a_h\cdot b(0)\quad =0. \end{cases}$$

Identifying the left members of the h^{th} equations in the two systems (97) and (196), we obtain $a_h \cdot b(0) = a_h$. Since $a_h \neq 0$, this relation gives

(197) $$b(0) = 1.$$

Inserting $b(0) = 1$ in the system (196), the $(h-1)^{\text{th}}$ equations of the two systems considered give $a_h \cdot b(-1) = 0$, or $b(-1) = 0$. In the same way we obtain successively

(198) $$0 = b(-1) = b(-2) = \cdots = b(-h+1).$$

Now, if b_t also satisfied a difference equation of lower order than h, it would follow from (198) that $b(0) = 0$, which contradicts (197).

We conclude from the theorem just proved that none of the individual components will be identically vanishing when b_t is written on the form (33). Thus, according to section 6, the two series $\sum_{i=1}^{\infty} |b_i|$ and $\sum_{i=1}^{\infty} b_i^2$ will be convergent if, and only if, all roots of the equation (34) are lying within the boundary of the unit circle. This corollary is important to the following.

Recurring to the solution $\{\xi(t_0;\, t)\}$, let the case be considered when the variables $\eta(t)$ are independent, and have identical distribution functions. Then we obtain from (195)

$$D^2(\xi(t_0;\, t)) = D^2(\eta(t)) \cdot [1 + b_1^2 + b_2^2 + \cdots + b_{t-t_0}^2].$$

Since $D(\eta_1) > 0$, the process $\{\xi(t_0;\, t)\}$ thus will be evolutive if the series $\sum_{i=1}^{\infty} b_i^2$ is divergent. According to the above, divergence takes place if one or more of the roots of the characteristic equation are lying on the boundary of or outside the unit circle.

The process of linear autoregression is, by construction, a solution of a stochastical difference equation such that (A) the variable in the right member is purely random, and (B) all roots of the characteristic equation are of a modulus less than unity. Leaving aside the question of whether there are other equations with stationary solutions — it seems likely that certain equations involving a singular and stationary $\{\eta(t)\}$ and where all the roots of the characteristic equation are of modulus unity are satisfied by suitable stationary and singular processes $\{\xi(t)\}$ — this short introduction will be terminated by the proof of the following theorem, which shows that the condition (A) can be generalized in as much

as it is sufficient that the right member is stationary. In the language of the theory of oscillatory mechanisms, the theorem states that a mechanism whose intrinsic movements are damped will give rise to a stationary oscillation when influenced by external shocks of a stationary kind.

Theorem 9. Let (193) *be a stochastical difference equation such that all roots of its characteristic equation are of a modulus less than unity, let* $\{\eta(t)\}$ *be stationary and have a finite dispersion, and let the sequence* $b_1, b_2, \ldots$ *be given by* (96) *and* (97). *Then*

$$\lim_{n\to\infty} [\{\eta(t)\} + b_1\cdot\{\eta(t-1)\} + b_2\cdot\{\eta(t-2)\} + \cdots + b_n\cdot\{\eta(t-n)\}]$$

will exist, and form a stationary solution of the equation.

Observing that for any $m > 0$

$$D^2(b_n\cdot\eta(t-n) + b_{n+1}\cdot\eta(t-n-1) + \cdots + b_{n+m}\cdot\eta(t-n-m)) \leq$$
$$\leq (|b_n| + |b_{n+1}| + \cdots + |b_{n+m}|)^2\cdot D^2(\eta(t)) \to 0 \text{ as } n\to\infty,$$

we conclude in the first place that

$$\lim_{n\to\infty} [\eta(t) + b_1\cdot\eta(t-1) + \cdots + b_n\cdot\eta(t-n)]$$

will exist. Denoting the limit variable by $\xi(t)$, an application of theorem 1 shows that the variables $\xi(t)$ arrived at will constitute a stationary process $\{\xi(t)\}$. That this process satisfies (193), may be proved in the same way as the identity (100).

23. On the stationary processes with finite dispersion and with no singular component.

When writing a stationary process with finite dispersion on the canonical form (176) it may happen that the singular component is vanishing (cf. section 15 γ). In this section, we shall derive some general formulae concerning such processes. In the coming sections of the present chapter, we shall use these formulae, and the previous analysis of stochastical difference equations, for a detailed study of the processes of moving averages and of linear autoregression.

Let $\{\zeta(t)\}$ and $\{\eta(t)\}$ represent two stationary processes such that

$$(199) \qquad \zeta(t) = \eta(t) + b_1 \cdot \eta(t-1) + b_2 \cdot \eta(t-2) + \ldots.$$

$$(200) \qquad \eta(t) = \zeta(t) + a_1 \cdot \zeta(t-1) + a_2 \cdot \zeta(t-2) + \ldots.$$

where

(A) $\{\eta(t)\}$ is non-autocorrelated,

(B) $D(\eta(t)) > 0$ is finite,

(C) $E[\eta(t)] = 0$,

(D) the sum $\Sigma\, b_k^2$ is convergent.

Thanks to the convergence of $\Sigma\, b_k^2$, the formulae involving only the coefficients b_k will all have a real meaning. On the other hand, the coefficients a_k have been introduced in a purely formal way, and all questions concerning their existence will be left open for later treatment in connexion with the analysis of special cases. As a matter of fact, the expressions which involve the coefficients a_k will be used mainly as a formal comprehension of the special cases of moving averages and of linear autoregression.

Replacing t in (199) by $t-1$, $t-2, \ldots$, and inserting in (200), we obtain the following relations between the coefficients a_i and b_i,

$$(201) \qquad a_k + b_1 \cdot a_{k-1} + \cdots + b_{k-1} \cdot a_1 + b_k = 0, \; k = 1, 2, \ldots$$

If the set (a_i) is given, the set (b_i) thus will be uniquely determined, and vice versa. We obtain for the first few coefficients

$$(202) \begin{cases} a_1 = -b_1, \\ a_2 = -b_2 + b_1^2, \\ a_3 = -b_3 + 2\, b_1\, b_2 - b_1^3, \\ a_4 = -b_4 + 2\, b_1 b_3 - 3 b_1^2 b_2 + b_2^2 + b_1^4, \end{cases} \begin{cases} b_1 = -a_1, \\ b_2 = -a_2 + a_1^2, \\ b_3 = -a_3 + 2\, a_1\, a_2 - a_1^3, \\ b_4 = -a_4 + 2\, a_1\, a_3 - 3\, a_1^2 a_2 + a_2^2 + a_1^4. \end{cases}$$

Keeping the notations used in section 19, we write

$$(203) \qquad K^2 = 1 + b_1^2 + b_2^2 + \ldots,$$

and remark that the vanishing of the singular component implies that $K = 1/\varkappa$. Thus we have

$$(204) \qquad D(\zeta) = K \cdot D(\eta),$$

while the general formulae (163) and (177) reduce to

$$K \cdot r(\zeta(t+n);\ \eta(t)) = b_n,$$

(205) $$r_k = r_k(\zeta) = (b_k + b_1 \cdot b_{k+1} + b_2 \cdot b_{k+2} + \cdots)/K^2,$$

writing shortly r_k for $r_k(\zeta)$.

Next, we shall derive a fundamental set of relations between the sequences (a), (b) and (r). Multiplying the two identities

(206) $$\zeta(t+s) = \eta(t+s) + b_1 \cdot \eta(t+s-1) + b_2 \cdot \eta(t+s-2) + \cdots$$

(207) $$\zeta(t) + a_1 \cdot \zeta(t-1) + a_2 \cdot \zeta(t-2) + \cdots = \eta(t),$$

and forming the expectations of the resulting two members, we obtain in the case of a negative s in (206)

(208) $$r_k + a_1 \cdot r_{k-1} + \cdots + a_{k-1} \cdot r_1 + a_k + a_{k+1} \cdot r_1 + a_{k+2} \cdot r_2 + \cdots = 0,$$

for all $k > 0$. In the same way, taking $s \geq 0$, we get

(209) $$(1 + a_1 \cdot r_1 + a_2 \cdot r_2 + \cdots) \cdot D^2(\zeta) = D^2(\eta),$$

and

(210) $$r_k + a_1 \cdot r_{k+1} + a_2 \cdot r_{k+2} + \cdots = b_k/K^2, \qquad k \geq 0.$$

Using the relations (149)—(152), we shall next derive a set of forecast formulae. For this purpose, we shall consider the variable $\zeta(t+k)$ as conditioned by the development of the process up to the time point t inclusive. Writing

(211) $$Fc[\zeta(t+k)] = F_t[\zeta(t+k)],$$

where $(C) = (\zeta(t-k) = \zeta_{t-k},\ \ \eta(t-k) = \eta_{t-k};\quad k = 0, 1, 2, \ldots)$, we have first the formal relations

(212) $$\begin{cases} \zeta_{t-k} = \eta_{t-k} + b_1 \cdot \eta_{t-k-1} + b_2 \cdot \eta_{t-k-2} + \cdots; & k = 0, 1, 2, \ldots \\ \eta_{t-k} = \zeta_{t-k} + a_1 \cdot \zeta_{t-k-1} + a_2 \cdot \zeta_{t-k-2} + \cdots; & k = 0, 1, 2, \ldots \end{cases}$$

Since the variables $\eta(t)$ are uncorrelated, (150) gives the linear forecast

(213) $$F_t[\zeta(t+k)] = b_k \cdot \eta_t + b_{k+1} \cdot \eta_{t-1} + b_{k+2} \cdot \eta_{t-2} + \cdots; \quad k = 1, 2, \ldots$$

As verified below, we have further

$$214) \quad F_t[\zeta(t+k)] = -a_1 \cdot F_t[\zeta(t+k-1)] - a_2 \cdot F_t[\zeta(t+k-2)] - \\ - \cdots - a_{k-1} \cdot F_t[\zeta(t+1)] - a_k \cdot \zeta_t - a_{k+1} \cdot \zeta_{t-1} - a_{k+2} \cdot \zeta_{t-2} - \\ - \cdots; \quad k = 1, 2, \ldots$$

This relation, which makes possible a successive calculation of the inear forecasts F_t, reduces to (213) if every $F_t[\zeta(t+k-i)]$ is vritten on the form (214), and every ζ_{t-i} is expressed in the values η_{t-i} by means of (212). In fact, for $i \geq 0$ the coefficient of η_{t-i} hen becomes

$$-a_1 \cdot b_{k+i-1} - a_2 \cdot b_{k+i-2} - \cdots - a_{k+i-1} \cdot b_1 - a_{k+i},$$

nd according to (201) this expression equals b_{k+i}.

Alternatively, we may express the forecasts F_t in terms of the alues ζ_{t-i}. Writing

$$215) \quad F_t[\zeta(t+k)] = f_{k,0} \cdot \zeta_t + f_{k,1} \cdot \zeta_{t-1} + f_{k,2} \cdot \zeta_{t-2} + \cdots,$$

'e record in the first place

$$f_{1,i} = -a_i.$$

urther, inserting (215) in the left member of (214), and writing so the forecasts in the right member on the same form, we btain

$$16) \quad \begin{cases} f_{k,0} + a_1 \cdot f_{k-1,0} + a_2 \cdot f_{k-2,0} + \cdots + a_{k-1} \cdot f_{1,0} + a_k = 0, \\ f_{k,1} + a_1 \cdot f_{k-1,1} + a_2 \cdot f_{k-2,1} + \cdots + a_{k-1} \cdot f_{1,1} + a_{k+1} = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

hus, after having calculated the coefficients $f_{k-i,j}$ appearing in the recasts $F_t[\zeta(t+k-i)]$, the relations (216) yield the coefficients $_j$ necessary for computing $F_t[\zeta(t+k)]$ in terms of the ζ_{t-i}'s.

The relations (213)—(215) will be referred to as the *forecasting rmulae*. The sample series sections $(\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \ldots)$ or (and) $(\eta_t,$ $_{-1}, \eta_{t-2}, \ldots)$ being given, these formulae furnish the best linear recast as to the future development of the series, viz. in the llowing sense.

A particular sample series section, say $(C) = (\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \ldots)$ ing given, there exists a set of constants $f_{k,0}(C), f_{k,1}(C), f_{k,2}(C), \ldots$ nimizing the expectation

$$7) \quad E[[\zeta_C(t+k) - f_{k,0}(C) \cdot \zeta_t - f_{k,1}(C) \cdot \zeta_{t-1} - f_{k,2}(C) \cdot \zeta_{t-2} - \cdots]^2].$$

In general, the constants $f_{k,i}(C)$ will depend on (C), and differ from the $f_{k,i}$'s. Now, considering on the other hand the expectation

$$E\,[[\zeta_C(t+k) - f_{k,0}\cdot\zeta_t - f_{k,1}\cdot\zeta_{t-1} - f_{k,2}\cdot\zeta_{t-2} - \cdots]^2],$$

our $f_{k,i}$'s possess the property that the (weighted) average of this expression based on all sets (C) becomes a minimum. This minimum is

$$(218) \qquad (1 + b_1^2 + b_2^2 + \cdots + b_{k-1}^2)\cdot D^2(\eta),$$

a formula which shows clearly the scope of the forecast method under consideration. A forecast $F_t[\zeta(t+k)]$ over k time units is decreasing in efficiency as k increases. As $k \to \infty$, the expression (218) tends to $K^2\cdot D^2(\eta) = D^2(\zeta(t))$. In other words, for large k-values the forecast $F_t[\zeta(t+k)]$ is approximately of the same efficiency as the trivial forecast $E[\zeta(t+k)] = E[\zeta(t)] = 0$.

If $\{\eta(t)\}$ is purely random, it follows from (152) that we have

$$F_t[\zeta(t+k)] = E_C[\zeta(t+k)].$$

In this case, the coefficients $f_{k,i}(C)$ appearing in (217) are independent of (C),

$$f_{k,i}(C) = f_{k,i}.$$

It is seen that (213)—(215) then give for every (C) a forecast which is the best one according to the principle of least squares.[13]

24. On the process of linear autoregression. General developments.

As already mentioned, we shall in the present section investigate in some detail the process of linear autoregression as dealt with in the sections 15 δ and 22. Denoting the process by $\{\zeta(t)\}$, the defining relation will be written

$$(219) \qquad \{\zeta(t)\} + a_1\{\zeta(t-1)\} + \cdots + a_h\{\zeta(t-h)\} = \{\eta(t)\}.$$

It will be observed that the process $\{\eta(t)\}$ need not be purely random; the following developments are valid under the broader assumption that the stationary process $\{\eta(t)\}$ is non-autocorrelated. As before, we shall assume that $E[\eta(t)] = E[\zeta(t)] = 0$.

The formal developments given in the previous section are all valid in the present case. In fact, in the sequence a_i we have $a_n = 0$ for $n > h$, so the serial developments are only apparently infinite. In order to arrive at more precise knowledge, we shall next consider these developments in some detail.

As to the b_t-series, we already know from the analysis in section 22 of the general stochastical difference equation that b_t in the present case is of type (33), and that the oscillations are damped, and that the series does not satisfy any linear difference equation of lower order than h.

Formula (209) gives

$$(1 + a_1 r_1 + \cdots + a_h r_h) \cdot D^2(\xi) = D^2(\eta). \tag{220}$$

As to the autocorrelation coefficients, we obtain from (208) and (210) the following three groups of relations.

$$\tag{221} \begin{cases} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ r_k + a_1 \cdot r_{k-1} + a_2 \cdot r_{k-2} + \cdots + a_{h-1} \cdot r_{k-h+1} + a_h \cdot r_{k-h} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ r_{h+1} + a_1 \cdot r_h + a_2 \cdot r_{h-1} + \cdots + a_{h-2} \cdot r_3 + a_{h-1} \cdot r_2 + a_h \cdot r_1 = 0 \\ r_h + a_1 \cdot r_{h-1} + a_2 \cdot r_{h-2} + \cdots + a_{h-2} \cdot r_2 + a_{h-1} \cdot r_1 + a_h = 0 \end{cases}$$

$$\tag{222} \begin{cases} r_{h-1} + a_1 \cdot r_{h-2} + a_2 \cdot r_{h-3} + \cdots + a_{h-2} \cdot r_1 + a_{h-1} + a_h \cdot r_1 = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ r_1 + a_1 + a_2 \cdot r_1 + \cdots + a_{h-2} \cdot r_{h-3} + a_{h-1} \cdot r_{h-2} + a_h \cdot r_{h-1} = 0 \end{cases}$$

$$\tag{223} \begin{cases} 1 + a_1 \cdot r_1 + a_2 \cdot r_2 + \cdots + a_{h-2} \cdot r_{h-2} + a_{h-1} \cdot r_{h-1} + a_h \cdot r_h = 1/K^2 \\ r_1 + a_1 \cdot r_2 + a_2 \cdot r_3 + \cdots + a_{h-1} \cdot r_h + a_h \cdot r_{h+1} = b_1/K^2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ r_k + a_1 \cdot r_{k+1} + a_2 \cdot r_{k+2} + \cdots + a_{h-1} \cdot r_{h+k-1} + a_h \cdot r_{h+k} = b_k/K^2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

The first group is given in the paper of SIR G. WALKER (1931) already referred to. We quote from the same paper the observations that (32) constitutes a difference equation satisfied by the r_k-series for $k \geq h$, that this relation is the same as that satisfied by the b_k-series, and that both series thus present damped oscillations of type (33).* SIR G. WALKER mentions further that the observa-

* K. STUMPFF ((1936), p. 35) gives the relation $b_k = r_k$, which is correct only in case $h = 1$. STUMPFF's deduction of the expectation relations (127) is based on this error.

tion that the b_k-series satisfies the difference equation (32) has already been given by G. U. YULE (1927) for the special case $h=1$.

The second group (222) contains $h-1$ relations, and involves the autocorrelation coefficients $r_1, r_2, \ldots, r_{h-1}$. Now, we obtain the coefficients $r_1, \ldots, r_{h-1}$ directly in terms of the coefficients a_i by solving the system (222). This important fact may be regarded as a corollary to the following theorem. To see this we assume to the contrary that equations (222) were connected by a linear relation, say with coefficients $A_1, \ldots, A_{h-1}$. An inspection of (222) with regard paid to (224)—(225) then shows that the r_k would satisfy a difference equation of order $h-1$, and this would contradict the theorem.

Theorem 10. Let $\{\zeta(t)\}$ be a process of linear autoregression of order h. Then the autocorrelation coefficients of $\{\zeta(t)\}$ satisfy no difference equation of type (32) *of lower order than h.*

For the proof, let us write r_k on the form (33), and consider the values, say $r(k)$, taken on by this function for $k=0, -1, -2, \ldots$ According to (177), we have $m=0$. Next, comparing the relation

$$r_h + a_1 \cdot r_{h-1} + \cdots + a_{h-1} \cdot r_1 + a_h \cdot r(0) = 0$$

with the last equation in the system (221), we find $a_h \cdot r(0) = a_h$. Since $a_h \neq 0$, we conclude

$$r(0) = 1. \tag{224}$$

Comparing further the first equation in the system (222) with

$$r_{h-1} + a_1 \cdot r_{h-2} + \cdots + a_{h-2} \cdot r_1 + a_{h-1} \cdot r(0) + a_h \cdot r(-1) = 0,$$

and paying regard to (224), we obtain $r(-1) = r_1$. This procedure may be continued $h-1$ steps, which gives

$$r(-1) = r_1,\ r(-2) = r_2,\ \ldots,\ r(-h+1) = r_{h-1}. \tag{225}$$

Proceeding one step further, we arrive at the first equation in the system (223), and obtain $a_h \cdot r(-h) + \frac{1}{K^2} = a_h \cdot r_h$, from which we conclude

$$r(-h) \neq r_h. \tag{226}$$

Now, considering the function d_k defined for $k = 0, \pm 1, \pm 2, \ldots$ by

$$d_k = r(k) - r(-k), \tag{227}$$

the relations (224)—(226) yield

$$d_{-h} \neq 0; \quad d_{-h+1} = d_{-h+2} = \cdots = d_{-1} = d_0 = d_1 = \cdots = d_{h-1} = 0; \quad d_h \neq 0. \tag{228}$$

On the other hand, it follows from (227) that d_k satisfies a difference equation of type (32) of an order twice that of the equation satisfied by r_k. However, by the argument used in theorem 8 it follows from (228) that d_k cannot possibly satisfy an equation (32) of lower order than $2h$. Thus r_k will satisfy no equation of type (32) of lower order than h, a reflection which completes the proof.

Among the general developments in section 23 there remains for consideration the forecasting formulae. Since $a_k = 0$ for $k > h$, formula (214) shows that the forecasts $F_t[\zeta(t+1)]$, $F_t[\zeta(t+2)]$, ..., $F_t[\zeta(t+k)]$, ... concerning the time points $t+1$, $t+2$, ..., $t+k$, ..., forecasts based on the development up to the time point t, will satisfy the equation (32) in respect of k. Thus, the forecasts, too, will form a damped oscillation. Explaining in the terminology of the theory of oscillatory mechanisms, the forecast curve describes how the mechanism would develop out from the situation arrived at in the time point t if there were no external influence present in the future time points $t+1, t+2, \ldots$ As by hypothesis the intrinsic oscillations of the mechanism are damped, the forecast curve will also be damped, in full agreement with the above. Thus, $F_t[\zeta(t+k)] \to 0 = E[\zeta(t)]$ as $k \to \infty$, in agreement with the concluding remark of section 23.

Considering, finally, the relations (216), it will be observed that the coefficients $f_{k,i}$ for all i satisfy the difference equation (32) in respect of k.

Next, we shall illustrate some points of the general analysis in Chapter II by means of the process of linear autoregression.

Keeping in mind that in the present case $\sum_{k=0}^{\infty} |r_k|$ is convergent, we may apply the corollary to theorem 5 (see p. 69). We conclude that the generating function $W(x)$ for all x in $(0, \pi)$ has a bounded derivative given by (121). In order to transform this expression upon a finite form we write

$$G(x) = \sum_{k=0}^{\infty} r_k \cdot e^{ikx} = e^{isx} \cdot \sum_{t=-s}^{\infty} r_{t+s} \cdot e^{itx}.$$

Considering the identity

$$G(x) \cdot [e^{-ihx} + a_1 \cdot e^{-i(h-1)x} + \cdots + a_{h-1} \cdot e^{-ix} + a_h] =$$
$$= \sum_{t=-h}^{\infty} r_{t+h} \cdot e^{itx} + a_1 \cdot \sum_{t=-h+1}^{\infty} r_{t+h-1} \cdot e^{itx} + \cdots +$$
$$+ a_{h-1} \cdot \sum_{t=-1}^{\infty} r_{t+1} \cdot e^{itx} + a_h \cdot \sum_{t=0}^{\infty} r_t \cdot e^{itx},$$

and paying regard to the relations (221), the right member reduces to

$$e^{-ihx} + (a_1 + r_1) \cdot e^{-i(h-1)x} + (a_2 + a_1 r_1 + r_2) \cdot e^{-i(h-2)x} + \cdots +$$
$$+ (a_{h-1} + a_{h-2} r_1 + \cdots + r_{h-1}) \cdot e^{-ix}.$$

Thus we obtain $G(x)$ on the finite form

$$G(x) = \frac{1 + (a_1 + r_1) \cdot e^{ix} + \cdots + (a_{h-1} + a_{h-2} r_1 + \cdots + a_1 r_{h-2} + r_{h-1}) \cdot e^{i(h-1)x}}{1 + a_1 \cdot e^{ix} + a_2 \cdot e^{2ix} + \cdots + a_h \cdot e^{ihx}},$$

while $W'(x)$ is given by

$$W'(x) = G(x) + G(-x) - 1 = 2 \cdot R[G(x)] - 1, \tag{229}$$

where $R[G(x)]$ stands for the real part of $G(x)$.

Denoting by x_i the roots of (34), the roots of $1 + a_1 x + \cdots + a_h x^h = 0$ are seen to equal $1/x_i$. Since these roots are lying outside the periphery of the unit circle, the denominator of $G(x)$ is evidently non-vanishing, which is in full agreement with the earlier observation that $W'(x)$ is bounded. Now, paying regard to the relations (127) and (129), and summing up the main results, we obtain the following theorem.

Theorem 11. The generating function $W(x)$ of the autocorrelation coefficients in a process $\{\xi(t)\}$ of linear autoregression is absolutely continuous, and has a bounded derivative $W'(x)$ given by (229). *The expectation $E[C^2(n;\lambda)]$ of an arbitrary ordinate in the* Schuster *periodogram, as defined by* (126), *is given by*

$$E[C^2(n;\lambda)] = \frac{4D^2(\xi)}{n} \cdot W'(\lambda) + O(1/n).$$

It should be noticed that the expectation is of the same order of magnitude as in the case of a purely random series (cf. (37)). When extending the analysed series, the periodogram ordinates thus tend to vanish.

In view of the applications of the theory of linear autoregression, it is of fundamental importance to investigate the possibilities of drawing conclusions from a sample series $(\zeta_t, \zeta_{t-1}, \zeta_{t-2}, \ldots)$ upon the characteristic equation (191), or — interpreting in the language of the theory of oscillatory mechanisms — to investigate whether the past development of the mechanism as influenced by random external factors can give any information about its intrinsic oscillatory tendencies. In particular, is it possible to find out the periods of the intrinsic damped oscillations?

The previous analysis shows that the classical periodogram analysis is an inadequate method for the search of intrinsic oscillations — the longer the series analysed, the poorer the results. This conclusion holds both for the Schuster and the Whittaker periodograms since they are of equal efficiency (see section 8). In the illustrations given in the next section, these periodogram questions are dealt with in more detail.

As has been emphasized by G. U. Yule (1927) and Sir G. Walker (1931), an adequate tool for the search of the intrinsic properties of an oscillatory mechanism is yielded by the serial coefficients of the time series investigated. In fact, a serial coefficient $\bar{r}_k$ approximates the corresponding autocorrelation coefficient r_k, and we know that the graph of r_k presents exactly those damped oscillations which are characteristic of the mechanism considered — there is conformity in respect to both the frequencies of the individual components of the oscillations and their damping exponential factors. Since a periodogram analysis is concerned with only the intrinsic frequencies, it is of particular importance that the serial coefficients can give information even about the damping factors.

Having above derived expressions for the autocorrelation coefficients and other characteristics connected with a process of linear autoregression, we are, in view of the applications, confronted with problems of an inverse type. In particular, it is seen that if the autocorrelation coefficients of a process of autoregression are known, the coefficients (a) will be obtained from the system (221—222). After having derived the coefficients (a) we obtain the primary process $\{\eta(t)\}$ in terms of the process considered by means of form-

ula (200). Concluding that the inverse problem mentioned involves no difficulty in point of principle, reference is given to the next section and Chapter IV for illustrations.

A question now presenting itself concerns the reliability of the information yielded by the serial coefficients. Here we meet at once the same obstacles as in all significance problems concerning autocorrelated time series. In the first place, we notice that when forming the sampling variance of a serial coefficient, we arrive at a complicated expression involving i. a. an extensive sum of correlation coefficients between different serial coefficients. The difficulties of the problem having already been mentioned by G. U. YULE (1927), E. SLUTSKY (1934) has presented a large collection of formulae concerning the dispersion of various characteristics derived from sample series sections, formulae deduced under the assumption that the variables considered are normally distributed.

However, it should be observed that the relevant problem does not consist merely in calculating the variance or the distribution of a single autocorrelation coefficient. We have also to face the much deeper question concerning the reliability of the periodicities which present themselves in the graph of the serial coefficients. In view of the complications already occurring in sampling problems involving merely one individual serial coefficient, the possibility of arriving at a practicable, quantitative measure of significance in this connexion seems, at least for the moment, hopeless.[14]

There is also another essential difficulty. Correlation in time series, and correlation as considered in the classical applications, differ as to their quantitative significance (see H. WOLD (1936)). As a matter of fact, in the former case the correlation coefficients are, as a rule, quantitatively conditioned by the size of the statistical masses to which the coefficients refer. In order to give an example, we advance that certain business cycle data yield nice graphs of serial coefficients (see Chapter IV). For instance, the serial coefficients of the G. MYRDAL index 1830—1913 of the cost of living in Sweden show a clear damped oscillation (see e. g. fig. 14). Let it now be imagined that another index had been computed by the same method, covering the same period — 84 years —, but referring merely to a small part of Sweden. In the second case there would, of course, be a larger random element present in the index. A short reflection will show that the increase of the random element

is accompanied by a systematic tendency to a diminishing of the serial coefficients. — Consider, on the other hand, a classical application such as the correlation between cranial indices. Here the statistical units, the skulls, are uniquely determined, *unmodifiable* — in a material of say 84 skulls the correlation coefficient can be calculated in only one way. In other words, while we cannot possibly form a coefficient referring to a certain part of each of the 84 skulls, there is nothing illegitimate in the modifying of the 84 statistical units in the case of the time series correlation considered above. Referring to Appendix B of the 1st ed. of this book for further comment, we note by way of a general conclusion that in a theoretical autocorrelation model the size of the statistical mass must also be taken into consideration.

Summing up, the intricacy of the problems of significance, and the dependence of the correlation coefficients on the statistical mass, constitute two fundamental difficulties in quantitative autocorrelation analysis. The situation seems to justify the opinion that the utmost caution is necessary when drawing quantitative conclusions from observational time series — a hypothesis should not be considered safe unless corroborated by empirical series obtained from different and, if possible, independent statistical masses, and supported by aprioristic considerations independent of the statistical evidence.

The applications to observational data presented in Chapter IV are far from aiming at quantitative results, the purpose being more to exemplify the qualitative differences between the scheme of hidden periodicities and the schemes of linear regression. Since we attach no importance to the quantitative outcomes, the significance problems will not be entered upon in the present study.

25. On some special cases of linear autoregression.

Keeping the assumptions of the previous section, we shall in this section consider the special cases obtained when putting $h=1$ and $h=2$ in the general definition (219) of linear autoregression. The resulting formulae will be readily surveyable, and give rise to a few remarks of general scope. In a few instances the model series given in section 15 will be used for the illustrations.

Attaching the analysis to the formula (219), we take $h = 2$, and denote the roots of the characteristic equation (191) by p and q. Thus we have

$$(230) \quad a_1 = -(p+q),\ a_2 = p \cdot q,\ a_n = 0 \text{ for } n > 2; \quad |p| < 1,\ |q| < 1;$$
$$\zeta(t) - (p+q) \cdot \zeta(t-1) + pq \cdot \zeta(t-2) = \eta(t).$$

As the coefficients a_i must be real, we notice that either I. p and q are real, or II. $p = A + iB$, $q = A - iB$, where A and B represent real numbers such that

$$(231) \qquad A^2 + B^2 = |p|^2 = |q|^2 < 1.$$

We shall assume that $p \neq q$, gaining thereby the general solution of the difference equation (32) to be (cf. also p. 146)

$$(232) \qquad P_1 \cdot p^t + P_2 \cdot q^t,$$

where P_1 and P_2 are arbitrary. In case II, this expression may be written (cf. (33))

$$Q_1 \cdot C^t \cos \lambda_1 t + Q_2 \cdot C^t \sin \lambda_1 t.$$

where Q_1 and Q_2 are arbitrary, and

$$C = +\sqrt{A^2 + B^2}; \qquad \cos \lambda_1 = A/C,\ 0 < \lambda_1 < \pi.$$

Cases I and II will be dealt with separately.

I. *p and q are real.*

Inserting the general expression (232) for b_1 and b_2 in the system (97), and solving for the constants P_1 and P_2, we obtain readily

$$(233) \quad b_k = \frac{p}{p-q} \cdot p^k + \frac{q}{q-p} \cdot q^k = (p^{k+1} - q^{k+1})/(p-q),\ k \geq 0.$$

Next, insertion of this expression in (203) yields

$$K^2 = \frac{D^2(\zeta)}{D^2(\eta)} = \frac{1 + pq}{(1-p^2)(1-q^2)(1-pq)}.$$

The system (222) reduces to the single equation

$$(234) \qquad r_1 + a_1 + a_2 \cdot r_1 = 0.$$

Solving for r_1, observing that $r_0 = 1$ (cf. formula (224)), equalling these two cofficients to the general expression (232) for r_0 and r_1, and solving the linear system thus obtained for the constants P_1 and P_2, we get

$$(235) \qquad r_k = \frac{p(1-q^2)}{(p-q)(1+pq)} \cdot p^k + \frac{q(1-p^2)}{(q-p)(1+pq)} \cdot q^k, \; k \geq 0.$$

It is readily verified that (205) is satisfied. Using (230) and (235) in the formula (229), we find for the derivative of the generating function in the interval $(0, \pi)$

$$(236) \qquad W'(\lambda) = \frac{(1-p^2)(1-q^2)(1-pq)}{(1+pq)(1+p^2-2p\cos\lambda)(1+q^2-2q\cos\lambda)} =$$

$$= \frac{1}{K^2 \cdot (1+p^2-2p\cos\lambda)(1+q^2-2q\cos\lambda)}$$

and a short calculation will verify (121).

Theorem 11 gives for the expectation of an arbitrary ordinate in the SCHUSTER periodogram

$$E[C^2(n,\lambda)] = \frac{4D^2(\eta)}{n(1+p^2-2p\cos\lambda)(1+q^2-2q\cos\lambda)} + o(1/n).$$

The following special cases are instructive.

1) $q = 0$. In this case the relations reduce to

$$(237) \qquad \zeta(t) - p\,\zeta(t-1) = \eta(t),$$

$$(238) \qquad b_k = p^k, \qquad D^2(\zeta) = D^2(\eta)/(1-p^2), \qquad r_k = p^k, \; k \geq 0,$$

$$W'(\lambda) = \frac{1-p^2}{1+p^2-2p\cos\lambda} = \frac{D^2(\eta)}{D^2(\zeta)\cdot(1+p^2-2p\cos\lambda)};$$

$$(239) \qquad E[C^2(n,\lambda)] \cong \frac{4D^2(\eta)}{n\cdot(1+p^2-2p\cos\lambda)}.$$

These formulae, which cover the case $h = 1$, have been given earlier by SIR G. WALKER ((1931), p. 521).*

* SIR G. WALKER gives also the variance formula (220) for an arbitrary h (see his formula K).

2) $q = -p$. We obtain without difficulty[15]

$$\zeta(t) - p^2 \zeta(t-2) = \eta(t),$$

$$b_{2k} = r_{2k} = p^{2k}, \ b_{2k+1} = r_{2k+1} = 0, \ k \geq 0$$

$$D^2(\zeta) = D^2(\eta)/(1-p^4),$$

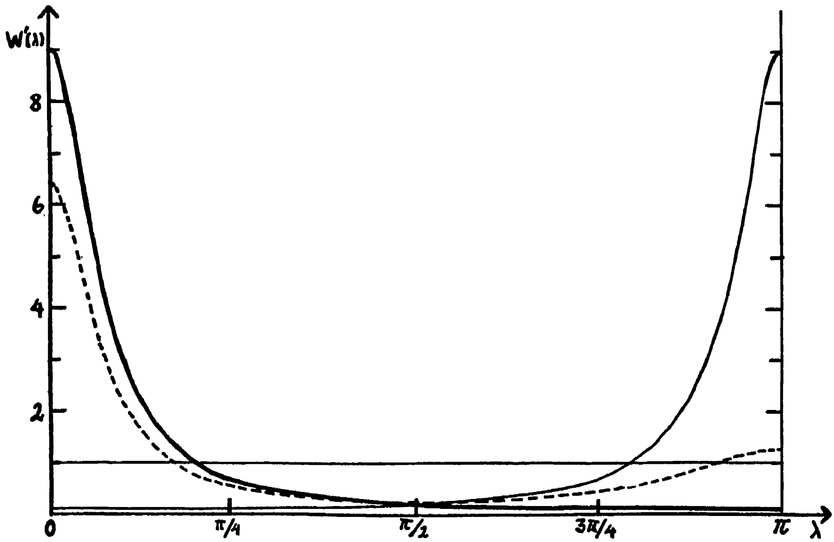

Fig. 2. Generating function derivatives obtained from formulae (236) *and* (239).

(240) $$W'(\lambda) = \frac{1-p^4}{(1+p^2)^2 - 4\,p^2 \cdot \cos^2 \lambda},$$

$$E[C^2(n, \lambda)] = \frac{4\,D^2(\eta)}{n \cdot [(1+p^2)^2 - 4\,p^2 \cdot \cos^2 \lambda]} + 0(1/n).$$

The graph above shows the curves $W'(\lambda)$ which belong to the processes defined by

a) formula (103), with $p = -{\cdot}8$, $q = 0$ (thin line),

b) » (104), with $p = {\cdot}8$, $q = 0$ (thick line),

c) $p = {\cdot}8$, $q = -{\cdot}6$ (broken line).

The graph contains, for comparison, the line $W'(\lambda) = 1$ which represents the derivative of the generating function in the case of a purely random process (cf. the remark attached to formula (131)).

II. *p and q are conjugate complex,*

(241) $$p = A + iB, \quad q = A - iB.$$

Since the above developments are perfectly general, we have now only to insert (241) in the formulae for b_k, r_k, etc. given under I. After elementary transformations we get for $k \geq 0$

$$\zeta(t) - 2A \cdot \zeta(t-1) + (A^2 + B^2) \cdot \zeta(t-2) = \eta(t),$$

(242) $$b_k = C^k \cos k\lambda_1 + \frac{A}{B} \cdot C^k \sin k\lambda_1,$$

(243) $$r_k = C^k \cos k\lambda_1 + \frac{A}{B} \cdot \frac{1 - C^2}{1 + C^2} C^k \sin k\lambda_1,$$

(244) $$D^2(\zeta) = \frac{1 + C^2}{(1 - C^2)(1 + C^4 - 2A^2 + 2B^2)} \cdot D^2(\eta).$$

Writing

$$J(\lambda) = 4\, C^2 \cdot \left[\frac{A(1 + C^2)}{2\, C^2} - \cos\lambda\right]^2 + \frac{B^2}{C^2}(1 - C^2)^2,$$

we get

(245) $$W'(\lambda) = \frac{D^2(\eta)}{D^2(\zeta) \cdot J(\lambda)}$$

and

$$E[C^2(n, \lambda)] = \frac{4\, D^2(\eta)}{n \cdot J(\lambda)} + o\,(1/n).$$

Possessing now a collection of formulae sufficient for our purposes, a partial check is obtained by observing that when putting $A = 0$ in the above expressions, we get the same result as when replacing p by $i \cdot p$ in the formulae I, 2.

Proceeding to a first application of the above developments, we observe that the formulae under II cover the case of an oscillatory mechanism whose intrinsic oscillations consist of a single damped harmonic with frequency λ_1 lying in the interval $0 < \lambda_1 < \pi$, and a damping factor C^t. Having earlier found that a periodogram analysis is ineffective in the case of linear autoregression, we are now in a position to prove the statement advanced in section 21 concerning the dangers of the periodogram method (see p. 95).

Speaking generally, the question is whether the expectation of the periodogram ordinate $C^2(n, \lambda)$ presents a maximum when $\lambda = \lambda_i$, i. e. when the abscissa of the periodogram equals a frequency characteristic to the intrinsic oscillations of the mechanism. Were the answer in the affirmative, we should — at least in principle — be able to use a periodogram construction for the discovery of the intrinsic frequencies λ_i. In fact, considering a sample series $(\zeta_{t-1}, \zeta_{t-2}, \ldots)$, making a sequence of periodograms on the basis of sections of type $(\zeta_{t-1}, \ldots, \zeta_{t-n})$, $(\zeta_{t-n-1}, \ldots, \zeta_{t-2n})$, ..., and forming for every abscissa λ the average of the corresponding sequence of periodogram ordinates, the resulting curve would present maxima in the points $\lambda = \lambda_i$ sought for. However, in order to prove that this way is barred — in point of principle, thus even if the statistical material were extensive enough for the construction of the required set of periodograms — we need only consider the case of a single intrinsic oscillation covered by the formulae (II) above. The expectation $E[C^2(n, \lambda)]$ being asymptotically proportional to the derivative $W'(\lambda)$, it will be sufficient to investigate the extremes of the latter function.

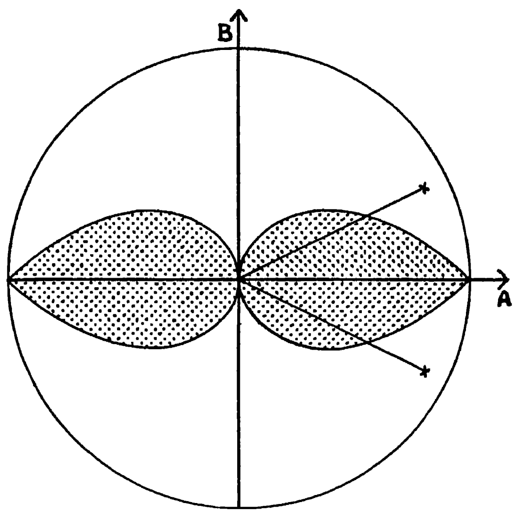

Fig. 3. Unit circle showing where $W'(\lambda)$ as given by (245) presents one or no maximum for $0 < \lambda < \pi$ (non-dotted and dotted domains respectively).

Evidently, $W'(\lambda)$ as given by (245) is maximized by those λ-values which minimize the auxiliary function $J(\lambda)$, and *vice versa*. The λ-values in question are seen to be

$$(246) \qquad \lambda = \text{arc cos}\, \frac{A(1 + C^2)}{2\,C^2}, \quad \lambda = 0, \quad \lambda = \pi.$$

The behaviour of $W'(\lambda)$ being different in the cases $|A| \gtrless 2\,C^2/(1 + C^2)$, the accompanying figure shows the part of the curve $|A| = 2\,C^2/(1 + C^2)$ lying in the unit circle $C = 1$. An analysis of the second derivative of $J(\lambda)$ gives the following results.

a) $|A| < 2C^2/(1+C^2)$. Referring to the figure, the roots $A \pm iB$ are in this case lying in the non-dotted part of the unit circle.

$W'(\lambda)$ presents one maximum in the interval $(0, \pi)$, viz. in the point λ defined by

$$\lambda = \text{arc cos } A(1 + C^2)/2C^2, \tag{247}$$

and two minima, viz. in the points $\lambda = 0$, and $\lambda = \pi$.

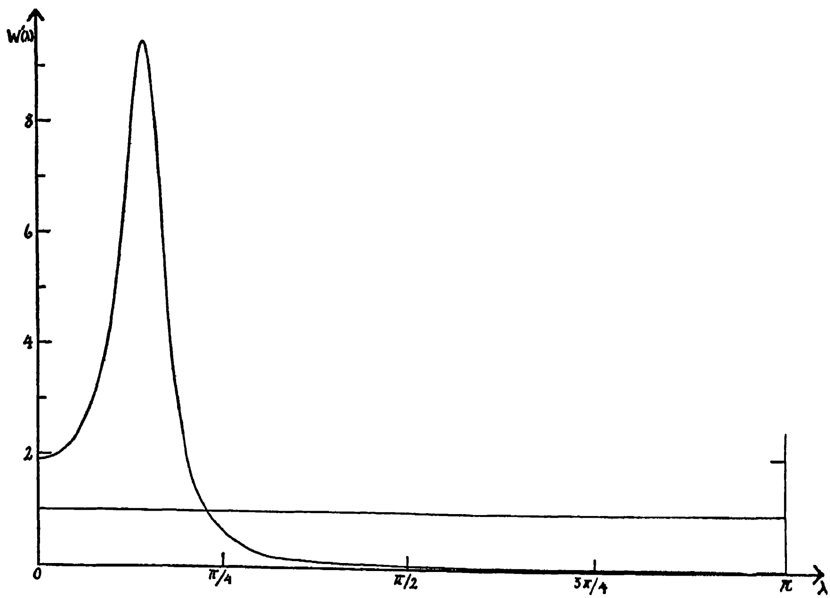

Fig. 4. Generating function derivative obtained from formula (245).

Illustration. Considering, for example, the process of type II obtained by taking $A = \cdot 8$, $B = \cdot 4$, we get $C^2 = \cdot 8$, $A < 2C^2/(1+C^2)$. In full agreement herewith, the point $(A, B) = (\cdot 8, \cdot 4)$, which is plotted in fig. 3, is lying in the non-dotted part of the unit circle. It follows from the above that the corresponding function $W'(\lambda)$ presents but one maximum in the interval $(0 \leq \lambda \leq \pi)$. This function $W'(\lambda)$ is shown in the figure above.

b) $|A| \geq 2C^2/(1 + C^2)$. The roots are lying in the dotted part of the unit circle.

$W'(\lambda)$ presents one maximum and one minimum, the former being attained for $\lambda = 0$ if $A > 0$ and for $\lambda = \pi$ if $A < 0$, the latter for $\lambda = \pi$ if $A > 0$ and for $\lambda = \pi$ if $A < 0$. The curves $W'(x)$ are here similar to those obtained in the case I, 1 (cf. fig. 2, unbroken lines).

Roughly speaking, if the roots $A \pm iB$ of the characteristic

equation of the oscillatory mechanism are lying close to the periphery of the unit circle, the maximizing λ-value (247) is seen to approximate the intrinsic frequency $\lambda_1 = \text{arc cos } A/C$. In other words, if the intrinsic oscillations are only slightly damped, the periodogram analysis suggested above will be able to discover the frequency of the intrinsic oscillation.

On the other hand, holding A/B fixed, and letting C leave the immediate neighbourhood of the periphery of the unit circle, the λ-values resulting from (247) will deviate more and more from $\lambda_1 = \text{arc cos } A/C$. In case $A > 0$, the value given by (247) is seen to be less than λ_1, while the reverse holds true if $A < 0$. Thus, the periodogram will show a tendency to over-estimate the intrinsic period if this is above 4 time units, and to under-estimate it if the period is lying between 2 and 4 units. As to periods below 2 time units, these require a finer equidistance (cf. section 4). — We conclude next that the bias will be the larger, the more heavily the intrinsic oscillation is damped, i. e. the smaller the damping factor C is. Further, excepting the cases of an intrinsic period equalling exactly 2 or 4 time units, and letting the damping factor pass below a well-defined limit, the periodogram will altogether cease from giving information about the intrinsic period.

The disturbing effect pointed out may be looked upon as caused by the external factors influencing the oscillatory mechanism. As by hypothesis these factors contain a random element, I propose the term *chance effect* for the bias in question.

The situation may be described by saying that the inference drawn from the characteristic equation of the intrinsic oscillations does not apply directly to the oscillations of the mechanism when influenced by random external factors (cf. p. 95). While the chance effect is easily surveyed in the case of one intrinsic damped harmonic, the state of things is far more complicated when the mechanism presents a tendency to composite oscillations. Having stated this, our analysis of the chance effect will be brought to an end by a few explicit illustrations.

Illustrations. We shall first consider the case of one intrinsic oscillation, with damping factor $C = \sqrt{\cdot 8} = \cdot 894$, and frequency $\lambda_1 = \text{arc tg } {}^1/_2 = 26^\circ, 56$. This case is elucidated by the diagrams 3 and 4. Inserting $A = \cdot 8$ and $C = \sqrt{\cdot 8}$ in (246), the λ-value maximizing $W'(\lambda)$ is found to be $\lambda = \text{arc cos } \cdot 9 = 25^\circ, 84$. The periodogram thus tends to deliver a period equalling $360/25 \cdot 84 = 13 \cdot 93$ time units, while the intrinsic period is $13 \cdot 55$ time units. In agreement with the general results concern-

ing the case of a non-composite intrinsic oscillation, the period is over-estimated. Although the damping factor is fairly large, just below unity, the bias is rather important.

Proceeding to some examples of periodogram ordinates derived from the model series given in section 15, let us first consider the process $\{\delta^{(1)}(t)\}$ as defined by (103). Continuing in the notations of the present section, we obtain by short calculations $h=1$, $p=-\,{}^{\cdot}8$, $D^2(\eta)={}^{\cdot}6$, $D^2(\zeta)=1{\cdot}667$. Speaking in the language of oscillatory mechanisms, we are concerned with a single intrinsic harmonic, the frequency, period, and damping factor of which are π, 2, and ˙8 respectively. An ordinary periodogram analysis gives negative results, the expectance being inversely proportional to the length of the series analyzed. However, we know from the general theory that the expectation of the periodogram ordinate in $\lambda=\pi$ is larger than the expectation in a purely random series. Taking $n=20$, and keeping in mind that we are dealing with the exceptional case $\lambda=\pi$, the latter is found to equal $1{\cdot}667/n={}^{\cdot}083$. The former, as given by the exact formula (127) is $E[C^2(20,\pi)]={}^{\cdot}59$. The approximate value obtained from (239) equals ˙75. Now, the 1000 elements in the model series $(\delta_t^{(1)})$ have been arranged in 50 sections, each containing 20 consecutive elements, and the periodogram ordinate $\overline{C}^2(20,\pi)$ has been computed for each section from formula (27). The average of the 50 ordinates thus obtained equals ˙56, a value not far from the the expectation ˙59.

Considering next the process $\{\delta^{(2)}(t)\}$ given by (104) we have $h=1$, $p={}^{\cdot}8$, $D^2(\eta)={}^{\cdot}2$, $D^2(\zeta)={}^{\cdot}556$. Referring again to Fig. 2, it is seen that for small frequencies the expectation is larger than in the case of a purely random series, but as before of the same order of magnitude in respect of n. Considering, e. g., the FOURIER coefficients (25) for $k=1$ in a sample series section of 20 elements, we have $n=20$, and $\lambda=2\pi/n=18^\circ$. Formulae (127) and (239) give respectively $E[C^2(20,\,18^\circ)]={}^{\cdot}36$ and $E[C^2(20,\,18^\circ)]\sim{}^{\cdot}34$, while the corresponding value in the purely random case is ˙11. On the other hand, operating on the model series $(\delta_t^{(2)})$ in the same way as before on $(\delta_t^{(1)})$, we have arrived at an average periodogram ordinate $\bar{E}[\bar{C}^2(20,\,18^\circ)]$ equalling ˙51. The rather substantial deviation from the expectation suggests that the periodogram ordinates are subject to a large dispersion. Actually, this suggestion seems to indicate how the matter stands. At any rate, the distributions of periodogram ordinates I have constructed on the basis of model time series have all presented a pronounced skewness, and a very large dispersion — often two or three of the largest sample values $\overline{C}^2(n,\lambda)$ constitute alone as much as 20 to 30 per cent of the sum of the 50 sample values in the material. An instructive example of this is given below.

Recurring to the series $(\delta_t^{(1)})$, formula (127) yields $E[C^2(20,\,18^\circ)]={}^{\cdot}046$, while (239) gives $E[C^2(20,\,18^\circ)]\sim{}^{\cdot}038$. A periodogram analysis as described in the previous illustration has given $\bar{E}[\overline{C}^2(20,18^\circ)]={}^{\cdot}058$. The figure below shows the distribution of the 50 averaged sample values of the periodogram ordinate dealt with. The figure contains the curve of summed relative frequencies ($\bar{F}$, thick lines), and a histogram showing the frequencies in the classes $0-{}^{\cdot}01$, ${}^{\cdot}01-{}^{\cdot}02$, etc ($f$, broken lines). The frequency curve graduating the histogram has been drawn by hand (broken curve). The distribution is seen to be very skew, and the histogram suggests no pronounced maximum for the frequency curve.

The above discussion is related to certain results obtained by E. SLUTSKY (1934) about sampling problems in periodogram analysis. It has long been recognized, of

course, that great caution is necessary when judging the reliability of a periodogram ordinate, especially if we have no prior knowledge about possible periods. There is further the difficulty that the largest periodogram ordinate has a larger expectation than a fixed ordinate, or a randomly chosen ordinate. A classic result by R. A. Fisher (1929) gives the distribution of the largest ordinate in the case of a purely random normal process.

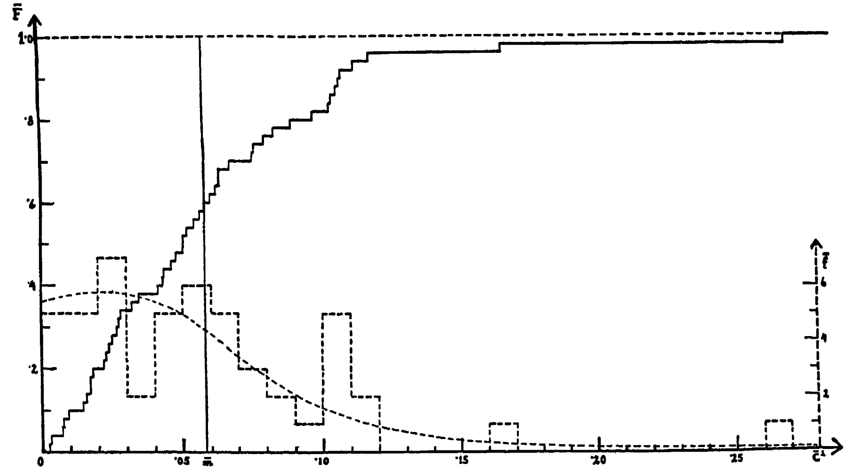

Fig. 5. Distribution of periodogram ordinates derived from the model series $\bar{\delta}_t^{(1)}$ (see table 3).

Next, we shall give a few applications of the inversion formula (200). Working on the model series given in section 15 δ, we shall exemplify the construction of the primary series from a given series of autoregression.

Illustrations. Considering the model series $(\bar{\delta}_t^{(1)})$ as defined by (103), we have $\bar{a}_t^{(2)} = \bar{\delta}_t^{(1)} + \cdot 8\, \bar{\delta}_{t-1}^{(1)}$ for $t = 1, 2, \ldots$ Inserting successively the $\bar{\delta}$-values given in table 3, we get $\bar{a}_1^{(2)} = 1\cdot 00$, $\bar{a}_2^{(2)} = \cdot 20 + \cdot 8 \times 1\cdot 00 = 1\cdot 00$, $\bar{a}_3^{(2)} = - \cdot 16 + \cdot 8 \times \cdot 20 = = \cdot 00$, etc., in full agreement with table 1. Thus, knowing the autoregression coefficient $a_1 = \cdot 8$, we can reconstruct the primary series of the autoregression series examined. In the same way, we can reconstruct without difficulty the primary series of the model series $(\bar{\delta}_t^{(2)})$ and $(\bar{\delta}_t^{(3)})$.

In view of the applications, an important problem is to find the coefficients (a) belonging to a given series of autoregression. This problem will be discussed in detail in section 32. For the moment we shall only remark that the relations (222) together with the last relation in (221) form a system of linear equations which will permit us to derive the coefficients $a_1, \ldots, a_h$ in terms of the autocorrelation coefficients $r_1, \ldots, r_h$. Thus, identifying the coefficients r_k with the serial coefficients $\bar{r}_k$ of the series examined, the linear system will give a set of approximate

coefficients $a_1, \ldots, a_h$. We can then derive as before the primary series which corresponds to the coefficients $a_1, \ldots, a_h$ arrived at.

We give below the first five serial coefficients of the model series ($\bar{\delta}$), and the corresponding autocorrelation coefficients as derived from formulae (238) and (243).

Table 5. Serial and autocorrelation coefficients of the model series ($\bar{\delta}$).

	$k =$	1	2	3	4	5
$\bar{\delta}t^{(1)}$	$\bar{r}_k =$	$-\cdot 786$	$\cdot 647$	$-\cdot 498$	$\cdot 375$	$-\cdot 291$
	$r_k =$	$-\cdot 800$	$\cdot 640$	$-\cdot 512$	$\cdot 410$	$-\cdot 328$
$\bar{\delta}t^{(2)}$	$\bar{r}_k =$	$\cdot 862$	$\cdot 700$	$\cdot 580$	$\cdot 482$	$\cdot 398$
	$r_k =$	$\cdot 800$	$\cdot 640$	$\cdot 512$	$\cdot 410$	$\cdot 328$
$\bar{\delta}t^{(3)}$	$\bar{r}_k =$	$\cdot 127$	$-\cdot 628$	$-\cdot 194$	$\cdot 397$	$\cdot 186$
	$r_k =$	$\cdot 121$	$-\cdot 626$	$-\cdot 204$	$\cdot 366$	$\cdot 206$

It is interesting to notice that although the model series consist of as many as 1000 elements each, the deviations between empirical and hypothetical correlation coefficients are rather substantial (cf. also p. 50 and p. 109).

Our analysis of the process of linear autoregression will be concluded by revealing a connexion with the »sinusoidal limit theorems» of E. Slutsky and V. Romanovsky touched upon in section 16.

Let

$$L(x) = x^2 - 2Ax + 1 = 0, \quad -1 < A < 1,$$

stand for the characteristic equation of a simple harmonic

$$P_1 \cos \lambda_1 t + P_2 \sin \lambda_1 t,$$

and let $\{\zeta^{(1)}(t)\}$, $\{\zeta^{(2)}(t)\}, \ldots$ represent a sequence of processes of linear autoregression defined by

$$\zeta^{(p)}(t) - 2A_p \cdot \zeta^{(p)}(t-1) + C_p^2 \cdot \zeta^{(p)}(t-2) = \eta^{(p)}(t).$$

Let it further be assumed that the processes have equal dispersion,

$$D(\zeta^{(p)}(t)) = \sigma,$$

and that

$$\lim_{p \to \infty} A_p = A, \quad \lim_{p \to \infty} C_p = 1.$$

cidating certain points of the autoregression analysis as presented in sections 19 and 20.

By definition, the general formula for a process $\{\zeta(t)\}$ of moving averages reads

$$(249) \qquad \{\zeta(t)\} = \{\eta(t)\} + b_1\{\eta(t-1)\} + \cdots + b_h\{\eta(t-h)\},$$

where $\{\eta(t)\}$ is purely random, and the sequence $(b) = (b_1, \ldots, b_h)$ is real. As before, we shall assume that $D(\eta)$ is finite, and that $E[\eta] = 0$. Thus (249) forms a special case of the variable defined by (199), and it follows further that the formal developments remain valid if $\{\eta(t)\}$ is non-autocorrelated.

In the process of linear autoregression, the autocorrelation coefficients and forecasting values were found to follow certain damped harmonics. In the present case only the first h elements in the sequences mentioned are different from zero. In fact, (204) and (205) yield for a process $\{\zeta(t)\}$ given by (249)

$$D^2(\zeta) = (1 + b_1^2 + b_2^2 + \cdots + b_h^2) \cdot D^2(\eta),$$

$$(250) \quad r_k(\zeta) = \begin{cases} (b_k + b_1 \cdot b_{k+1} + \cdots + b_h \cdot b_{h-k})/(1 + b_1^2 + \cdots + b_h^2) & \text{for } k \leq h, \\ 0 & \text{for } k > h, \end{cases}$$

where $k \geq 0$ and $b_0 = 1$, formulae given by Professor H. CRAMÉR in his 1933 Course. Next, (213) gives

$$F_t[\zeta(t+k)] = \begin{cases} b_k \cdot \eta_t + b_{k+1} \cdot \eta_{t-1} + \cdots + b_h \cdot \eta_{t-h+k} & \text{for } 0 \leq k \leq h, \\ 0 & \text{for } k > h, \end{cases}$$

where the forecast is based upon the condition

$$(C) = (\eta(t) = \eta_t, \eta(t-1) = \eta_{t-1}, \ldots, \eta(t-h+1) = \eta_{t-h+1}).$$

Inserting (250) in the general formula (116) for the generating function $W(x)$ of the autocorrelation coefficients in a stationary process, we conclude that $W(x) - x$ reduces to a finite trigonometrical sum. Thus, in the present case $W'(x)$ exists, and is like $W(x)$ uniformly bounded. Consequently, replacing »linear autoregression» by »moving averages» in theorem 11, we get a corresponding theorem covering the present case. We conclude, i. a., that periodogram analysis is an inadequate method of research in the case of moving averages also — the expectation of an arbitrary periodogram ordinate is inversely proportional to the length of the series under analysis. — Formula (256) gives $W'(x)$ explicitly.

In view of the applications, the following questions call for investigation (cf. the illustrations on p. 119). The autocorrelation coefficients r_k of a process of moving averages being given, is it possible to derive the primary process and the coefficients (b) of the moving average? In particular, does there exist a relation of type (200) yielding the primary process? If the answer to the second question is in the affirmative, how can the coefficients (a) be obtained? The first of these problems was formulated by Professor H. CRAMÉR in his 1933 Course.

It is suitable to start the analysis with the second question. Denoting the process considered by $\{\zeta(t)\}$, we shall, until further notice, assume that $\{\zeta(t)\}$ is given by (249).

In the first place we observe that theorem 9 covers the case when all the roots of the equation

$$x^h + b_1 x^{h-1} + \cdots + b_{h-1} x + b_h = 0 \tag{251}$$

are of modulus less than unity. The theorem states that if this condition is satisfied, an infinite sequence $(a) = (a_1, a_2, \ldots)$ such that (200) holds is given by the system (97) and the difference relations obtained for a_i when replacing the a_i's by the b_i's in (96). These relations constitute a difference equation of order h satisfied by the sequence (a), and it should further be observed that the sequence (a) arrived at forms a composed damped harmonic.

We have found that the condition attached to the roots of (251) will secure the representation (200). Under these circumstances the relations (208) and (210) derived from (206) and (207) must hold true. Explicitly, these read in the present case as follows:

$$\begin{aligned} a_{2h+k} r_h + a_{2h+k-1} r_{h-1} + \cdots + a_{h+k+1} r_1 + a_{h+k} + \\ + a_{h+k-1} r_1 + \cdots + a_{k+1} r_{h-1} + a_k r_h = 0, \end{aligned} \tag{252}$$

$$\begin{cases} a_{2h} r_h + a_{2h-1} r_{h-1} + \cdots + a_{h+1} r_1 + a_h + \\ \qquad + a_{h-1} r_1 + \cdots + a_1 r_{h-1} + r_h = 0 \\ a_{2h-1} r_h + a_{2h-2} r_{h-1} + \cdots + a_h r_1 + a_{h-1} + \\ \qquad + a_{h-2} r_1 + \cdots + a_1 r_{h-2} + r_{h-1} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{h+1} r_h + a_h r_{h-1} + \cdots + a_2 r_1 + a_1 + r_1 = 0 \end{cases} \tag{253}$$

$$4)\quad \begin{cases} a_h r_h + a_{h-1} r_{h-1} + \cdots + a_1 r_1 + 1 = 1/K^2 \\ a_{h-1} r_h + \cdots + a_1 r_2 + r_1 = b_1/K^2 \\ \cdots\cdots\cdots\cdots\cdots \\ a_2 r_h + a_1 r_{h-1} + r_{h-2} = b_{h-2}/K^2 \\ a_1 r_h + r_{h-1} = b_{h-1}/K^2 \\ r_h = b_h/K^2. \end{cases}$$

Comparing the above coefficients a_i and the coefficients $a_{i\,n} = - a(i,n)$ lded by an autoregression analysis of $\{\zeta(t)\}$ as described in section , and letting $\varepsilon > 0$ be arbitrarily given, it follows that there exists n such that

$$+ \varepsilon) D^2(\zeta) \geqq D^2(\zeta(t) + a_1 \zeta(t-1) + \cdots + a_n \zeta(t-n)) \geqq$$
$$\geqq D^2(\zeta(t) + a_{1\,n} \zeta(t-1) + \cdots + a_{n\,n} \zeta(t-n)) \geqq \varkappa^2 \cdot D^2(\zeta),$$

second inequality resulting from the definition of the coeffi- nts $a_{i\,n}$, and the third being similar to (182). Inserting (249), elementary transformation yields

$$\geq (a_1 + b_1)^2 + (a_2 + a_1 b_1 + b_2)^2 + \cdots +$$
$$+ (a_n + a_{n-1} b_1 + \cdots + a_{n-h} b_h)^2 + \cdots + (a_n b_h)^2 \geq$$
$$(a_{1\,n} + b_1)^2 + (a_{2\,n} + a_{1\,n} b_1 + b_2)^2 + \cdots +$$
$$+ (a_{n\,n} + a_{n-1,\,n} b_1 + \cdots + a_{n-h,\,n} b_h)^2 + \cdots + (a_{n\,n} b_h)^2 \geq 0.$$

ncluding that $|a_{i\,n} - a_i| < c_i \cdot \sqrt{\varepsilon}$, where c_i is independent of n, follows that $a_{i\,n} \to a_i$ as $n \to \infty$. Moreover, forming the difference the residuals based on $(a_1, \ldots, a_n)$ and $(a_{1\,n}, \ldots, a_{n\,n})$, a simple plication of SCHWARZ' inequality shows that the non-autocorrelated ocess yielded by the autoregression analysis is identical with the mary process.

Considering the singular case when (251) presents at least one root on the bound- of the unit circle, we conclude from theorem 8 that the series $\Sigma\, a_i^2$ will be ergent, and that no representation of type (200) will exist. By elementary trans- nations we shall show that this obstacle to calculating the primary process may removed by means of a limit passage.

Let (249) represent a moving average, x_n the roots of the characteristic equation 1), and let $|x_n| \leq 1$. Introducing auxiliary averages $\{\xi^{(i)}(t)\}$ defined by

$$\xi^{(i)}(t) = \eta(t) + B_1^{(i)} \cdot \eta(t-1) + \cdots + B_h^{(i)} \cdot \eta(t-h), \qquad i = 1, 2, \ldots,$$

let the symbols referring to $\{\xi^{(i)}(t)\}$ be marked by (i). Let $x_n^{(i)} = x_n$ when $|x_n| < 1$, and let $x_n - x_n^{(i)} = \varDelta_i$, $|x_n^{(i)}| = 1 - \varepsilon_i$ when $|x_n| = 1$, where $0 < \varepsilon_i = |\varDelta_i| \to 0$ as $i \to \infty$.

By construction, the inversion formula (200) applies to $\{\xi^{(i)}(t)\}$,

$$\eta(t) = \xi^{(i)}(t) + a_1^{(i)} \cdot \xi^{(i)}(t-1) + a_2^{(i)} \cdot \xi^{(i)}(t-2) + \cdots.$$

The limit relation desired is based on the coefficients $a_t^{(i)}$ and reads

$$(255) \qquad \{\eta(t)\} = \lim_{i\to\infty} [\{\zeta(t)\} + a_1^{(i)} \cdot \{\zeta(t-1)\} + a_2^{(i)} \cdot \{\zeta(t-2)\} + \cdots].$$

Writing $\{\eta(t, i)\}$ for the process under the limes sign, it is sufficient to prove that $D_i^2 = D^2(\eta(t,i) - \eta(t))$ is of the same order of magnitude as ε_i (see p. 40).

Paying regard to (198), we get $D_i^2 = \sum_{t=1}^{\infty} (L[a_t^{(i)}])^2$, where

$$L[x(t)] = x(t) + b_1 \cdot x(t-1) + \cdots + b_{h-1} \cdot x(t-h+1) + b_h \cdot x(t-h).$$

Write next $a_t^{(i)}$ on the form $\Sigma = \sum_n H^{(k_n-1)}(t) \cdot (x_n^{(i)})^t$, where $H^{(k)}$ is a polynomial of order k, and k_n the multiplicity of the root x_n (cf. (33)). Inserting $a_t^{(i)}$ in D_i^2, the terms in Σ of type $t^k \cdot (x_n^{(i)})^t$, where $x_n^{(i)} = x_n$, will cancel out because they satisfy the relations $L[t^k \cdot (x_n^{(i)})^t] = 0$. It remains to estimate the terms involving $t^k \cdot (x_n^{(i)})^t$, where $x_n^{(i)} = x_n - \varDelta_i$. According to the inequality of SCHWARZ it is evidently sufficient to verify that, for any of the k-values in question, $S = \sum_{t=1}^{\infty} (L[t^k \cdot (x_n - \varDelta_i)^t])^2 \to 0$ as $i \to \infty$.

To prove this, remove the factors $(x_n - \varDelta_i)^{t-h}$ from $L[t^k \cdot (x_n - \varDelta_i)^t]$, and develop the remaining terms $(x_n - \varDelta_i)^s$ according to the binomial theorem. Then we get

$$S = \sum_{t=1}^{\infty} (x_n - \varDelta_i)^{2t-2h} \cdot [P(t,k) + P_1(t,k) \cdot \varDelta_i + \cdots + P_k(t,k) \cdot \varDelta_i^k]^2,$$

where

$$P(t, k) = x_n^{-t+h} \cdot L[t^k \cdot x_n^t] = x_n^h \cdot t^k + b_1 \cdot x_n^{h-1} \cdot (t-1)^k + \cdots + b_h \cdot (t-h)^k.$$

Evidently $P(t,k) = 0$ for all t. Disregarding constant factors, we have further $P_1(t,k) = h \cdot x_n^h \cdot t^k + (h-1) b_1 \cdot x_n^{h-1} (t-1)^k + \cdots + b_{h-1} \cdot x_n \cdot (t-h+1)^k$, $P_2(t,k) = h(h-1) \cdot x_n^h \cdot t^k + (h-1)(h-2) b_1 \cdot x_n^{h-1} \cdot (t-1)^k + \cdots + 2 b_{h-2} \cdot x_n^2 \cdot (t-h+2)^k$, etc.

Paying regard to the identities $P(t,k) \equiv P(t,k-1) \equiv \cdots \equiv P(t,0) \equiv 0$, we get without difficulty $P(t+1, k+1) - P(t, k+1) = 0$ for all t. Hence $P(t, k+1)$ must reduce to a constant, say c_0. In the same way we get $P(t+1, k+2) - P(t, k+2) = c_1 \cdot P(t, k+1)$, which shows that $P(t, k+2)$ is linear in respect to t, and by induction we find that $P(t, k+s)$ is a polynomial in t of order $s-1$.

Elementary transformations show that $c_0 = P(t, k+1) - (t-h) \cdot P(t, k) = P_1(t, k)$ for all t. Similarly, $P(t, k+2) - (t-h+1) \cdot P_1(t, k) = P_2(t, k)$ is linear in respect of t, and repeating the procedure we find that $P_s(t, k)$ is of order $s-1$.

We conclude that S is linearly composed of a finite number of terms of type $S_k = \Delta_i^{k+2} \cdot \sum_{t=1}^{\infty} t^k \cdot (x_n - \Delta_i)^{2t}$, where $k \geq 0$. Hence $|S_k| \leq \varepsilon_i^{k+2} \cdot \sum_{t=1}^{\infty} t^k \cdot (1-\varepsilon_i)^{2t} \leq$ $\leq A \cdot \varepsilon_i^{k+2}/[1-(1-\varepsilon_i)^2]^{k+1}$, where A is independent of t. It is seen that for any finite $k \geq 0$ the term $|S_k|$ is of an order $\leq \varepsilon_i$, which completes the proof.

The above analysis has shown that if (251) has no root x_k falling outside the unit circle, certain well-defined linear operations on the moving average given by (249) will yield the primary process $\{\eta(t)\}$. If $|x_k| \leq 1$ for all k, the sequence (b) and the process $\{\zeta(t)\}$ will be termed *regular*.

Turning next to the generating function $W(x)$ defined by (121), let (249) be an arbitrary process of moving averages, and let its autocorrelation coefficients be represented by r_k. Paying regard to (250), we obtain the fundamental identity

$$(256)\quad \frac{1}{K^2}(x^h + b_1 x^{h-1} + \cdots + b_{h-1}x + b_h)(b_h x^h + b_{h-1} x^{h-1} + \cdots + b_1 x + 1) =$$
$$= r_h x^{2h} + r_{h-1} x^{2h-1} + \cdots + r_1 x^{h+1} + x^h + r_1 x^{h-1} + \cdots + r_{h-1} x + r_h.$$

Replacing x by e^{ix}, formula (121) shows that we get $e^{ihx} \cdot W'(x)$.

Until further notice, we shall again assume that (249) is regular. Denoting as before the zeros of the first factor by x_k, it is seen that the zeros of the second factor are given by $1/x_k$. Consequently, the zeros of the right member may be denoted $x_1, \ldots, x_{2h-1}, x_{2h}$, where

$$x_k = 1/x_{2h+1-k}, \quad 0 < |x_1| \leq |x_2| \leq \cdots \leq |x_h| \leq 1 \leq |x_{h+1}| \leq \cdots \leq |x_{2h}|.$$

Further, it follows that if there exists another sequence, say $(1, b_1^{(i)}, \ldots, b_h^{(i)})$, such that the corresponding moving average will have autocorrelation coefficients coinciding with those of (249), then one zero of the polynomial $x^h + b_1^{(i)} x^{h-1} + \cdots + b_{h-1}^{(i)} x + b_h^{(i)}$ will equal either x_1 or $1/x_1$, another either x_2 or $1/x_2$, etc. Evidently, there are at most 2^h real sequences of this type, say $(b_k^{(0)}), \ldots, (b_k^{(s)})$. Letting $(b_k^{(0)})$ represent the regular sequence started from, a short reflection reveals that all the other sequences in the group are non-regular.

If in this way a group $(b_k^{(i)})$ of sequences is attached to every regular sequence $(b_k^{(0)})$, it is clear that an arbitrary sequence $(b_k^*) =$

$=(1, b_1^*, \ldots, b_h^*)$ will belong to one, and only one, of these groups. It should further be observed that this group may be constructed with the use of only the sequence (b_k^*). Similarly, a group $(b_k^{(i)})$ will evidently be uniquely determined by the corresponding sequence of autocorrelation coefficients.

Thus prepared, let $\{\eta(t)\}$ be a purely random process, and let $(b_k^{(i)})$ be a group of finite sequences as defined above. Writing

$$(257) \qquad [K^{(i)}]^2 = 1 + [b_1^{(i)}]^2 + [b_2^{(i)}]^2 + \cdots + [b_h^{(i)}]^2,$$

let, correspondingly, a group $(\zeta^{(i)}) = (\{\zeta^{(i)}(t; \eta)\})$ of moving averages be defined by

$$(258) \qquad \zeta^{(i)}(t; \eta) = \frac{K^{(0)}}{K^{(i)}} [\eta(t) + b_1^{(i)} \eta(t-1) + \cdots + b_h^{(i)} \eta(t-h)].$$

Marking the symbols referring to different processes in a group by corresponding indices, it follows from the construction of the group $(\zeta^{(i)})$ that

$$(259) \qquad D(\zeta^{(i)}) = D(\zeta^{(j)}); \quad r_k^{(i)} = r_k^{(j)}, \quad k = 0, \pm 1, \pm 2, \ldots.$$

Further, the group will contain one, and only one regular process, viz. $\{\zeta^{(0)}(t)\}$. This will be assumed to be given by (249), and be alternatively denoted by $\{\zeta(t)\}$.

If all roots of (251) are falling on the periphery of the unit circle, the group $(\zeta^{(i)})$ will evidently contain only the process $\{\zeta^{(0)}(t)\} = \{\zeta(t)\}$. Otherwise the group will include more than one process, at most 2^h in number. Furthermore, it should be observed that an equivalent construction of the group $(\zeta^{(i)})$ is possible on the basis of the primary process $\{\eta(t)\}$ and the characteristics (259) common to the processes in the group.

Referring to the autoregression analysis as set forth in section 19, it is seen that the coefficients in the formulae for the residuals $\eta(t; n)$ involve only the autocorrelation coefficients and the dispersion of the process under analysis. Let this fact be combined with the above observation that the limit residual $\lim_{n \to \infty} \eta(t; n)$ of a regular process of moving averages may be obtained directly, viz. either from (200) or from (255). We conclude that the autoregression residuals, say $\{\eta^{(i)}(t)\}$, of the non-regular processes in a group $(\{\zeta^{(i)}(t; \eta)\})$ will be given by corresponding linear operations, and that these

expressions will involve exactly the same coefficients (a) as in the case of the regular average.

Considering in the first place the case when all the roots of the equation (251) obtained from the regular process (249) in the group are falling in the interior of the unit circle, the above argument yields

$$(260)\quad \frac{K^{(i)}}{K^{(0)}}\cdot\eta^{(i)}(t)=\eta(t)+(a_1+b_1^{(i)})\cdot\eta(t-1)+(a_2+a_1b_1^{(i)}+b_2^{(i)})\cdot\eta(t-2)+\cdots+ \\ +(a_h+a_{h-1}b_1^{(i)}+\cdots+b_h^{(i)})\cdot\eta(t-h)+ \\ +(a_{h+1}+a_hb_1^{(i)}+\cdots+a_1b_h^{(i)})\cdot\eta(t-h-1)+\cdots.$$

Since $(b^{(i)})\neq(b)$, the residual $\eta^{(i)}(t)$ cannot reduce to $\eta(t)$. Further, were $\eta^{(i)}(t)$ a finite moving average, it would follow that $r_k(\eta^{(i)})\neq 0$ for some $k>0$. As $\eta^{(i)}$ is non-autocorrelated this is impossible, so $\eta^{(i)}$ as given by (260) must be an infinite moving average (cf. section 15 γ). Now, paying regard to the relations (250) and (252—253), an elementary transformation will verify that $r_k(\eta^{(i)})=0$ for $k\neq 0$.

On the other hand, if at least one root of (251) is lying on the periphery of the unit circle, the representation (255) gives

$$\{\eta^{(i)}(t)\}=\lim_{k\to\infty}\left[\{\zeta^{(i)}(t)\}+a_1^{(k)}\{\zeta^{(i)}(t-1)\}+a_2^{(k)}\{\zeta^{(i)}(t-2)\}+\cdots\right].$$

By construction, the coefficients $B_n^{(k)}$ connected with (255) are such that $\lim_{k\to\infty}B_n^{(k)}=b_n^{(0)}=b_n$. Keeping this in mind, we conclude from (96) and (97) that $\lim_{k\to\infty}a_n^{(k)}=a_n$. Now, the relation (255) implies that we may express the ζ's in terms of the η's, and that the resulting sum, say $\Sigma\{\eta(t-k)\}\cdot c_k^{(i)}$, has a limit equalling the sum of the limits $\{\eta(t-k)\}\cdot\lim c_k^{(i)}$. Since $\{\zeta^{(i)}\}$ and $\{\zeta\}$ have identical autocorrelation properties, we may in the above relation perform the same procedure on $\{\zeta^{(i)}\}$. It follows that the representation (260) holds even in this case, without a limit passage being required.

Having now illustrated theorem 6 by means of a process of moving averages, the representation secured by theorem 7 is readily obtained directly. In fact, since the coefficients (b) in the canonical formula in theorem 7 are derived solely from the dispersion and the autocorrelation coefficients of the process considered, we conclude that they must be identical for all processes in a group $(\zeta^{(i)})$ as defined above. Thus we have (cf. also (258))

(261) $\{\zeta^{(i)}(t)\} = \{\eta^{(i)}(t)\} + b_1\{\eta^{(i)}(t-1)\} + \cdots + b_h\{\eta^{(i)}(t-h)\}$

for all processes in the group $(\zeta^{(i)})$. It should be noticed that (254) gives the coefficients (b) in terms of the sequences (a) and (r) characteristic to the group.

Summing up the main results, the analysis gives the following answers to the questions set forth on p. 123. The dispersion and the autocorrelation coefficients of a process of moving averages being given, there will in general exist a well-defined group of moving averages with the same characteristics, and with the same primary process. These moving averages are limited in number, and it is possible to construct the corresponding sequences of coefficients (b) by means of the characteristics prescribed. Alternatively, if one sequence (b) in a group is known, the others are uniquely determined. — Among the moving averages in a group it is possible to distinguish one, the regular process, which alone has the property that the primary process will be given either by a relation (200) or by a limit relation of the same type. The coefficients (a) in these representations are uniquely determined by the coefficients (b) of the regular process. — For the non-regular processes in a group, there exists no relation of type (200) yielding the primary process. However, inserting a non-regular process $\{\zeta^{(i)}(t)\}$ in the representation of the primary process in terms of the regular process belonging to the same group, we get the non-autocorrelated residual of $\{\zeta^{(i)}(t)\}$ secured by theorem 6. According to theorem 7, the process $\{\zeta^{(i)}(t)\}$ may be looked upon as a moving average of its residual. The coefficients of this average are nothing else than the coefficients (b) of the regular process in the group around $\{\zeta^{(i)}(t)\}$.

In connexion with the applications in section 31, we shall derive a linear relation of more general type than (200), which yields the primary process in terms of a non-regular moving average.

The autoregression analysis has given us no tool for distinguishing between the different moving averages in a group. If more precise information is required, other methods have to be applied, e.g. an analysis of conditioned variables (cf. p. 164). Such lines of research falling outside of the scope of the present study, this section will be terminated by some explicit examples of the previous analysis.

Illustrations. 1) Let the autocorrelation coefficients of a moving average be given by $r_1 = -\cdot5$, $r_n = 0$ for $n > 1$.

The relation (256) reads in this case

$$\cdot 5\,(x-1)\,(-x+1) = -\cdot 5\,x^2 + x - \cdot 5.$$

We conclude that the group $(b^{(i)})$ contains only one sequence, viz. (1, —1). Thus, the general formula of type (249) for a moving average with the given autocorrelation coefficients reads

$$\{\zeta(t)\} = \{\eta(t)\} - \{\eta(t-1)\}. \tag{262}$$

The equation (251) has only one root, $x_1 = 1$, and this falls on the unit circle. Hence, in order to express η in terms of ζ, we have to apply the limit relation (255). Taking, for instance, $x_1^{(k)} = 1 - 10^{-k}$, we get

$$\{\eta(t)\} = \lim_{k \to \infty} \left[\{\zeta(t)\} + (1-10^{-k}) \cdot \{\zeta(t-1)\} + (1-10^{-k})^2 \cdot \{\zeta(t-2)\} + \cdots\right]. \tag{263}$$

Table 2 contains a model series section of a process of type (262). Taking $k=1$, and applying (263) to the last element but one in this section, it is seen that we get the following approximation to the last element in table 1, (2),

$$-2 + 2\,(\cdot 9) - (\cdot 9)^3 + (\cdot 9)^4 - (\cdot 9)^5 - (\cdot 9)^6 + (\cdot 9)^7 + (\cdot 9)^9 + \cdots.$$

A computation of this sum has given $-\cdot 97$, which is fairly close to the exact value, i.e. -1. With any prescribed accuracy, it is possible to reconstruct in this way the series $(\alpha_t^{(2)})$ on the basis of a sufficiently long series (β_t).

2) Let the coefficients (b) of a moving average (249) be given by (1, 2). Forming the relation (256), we get

$$\cdot 2\,(x+2)(2x+1) = \cdot 4\,x^2 + x + \cdot 4,$$

and conclude that $r_1 = \cdot 4$, $r_n = 0$ for $n > 1$. The group $(b^{(i)})$ is seen to consist of $(1, \cdot 5)$ and (1, 2), the former sequence being the regular one. Now, the system (97) gives $a_1 = -\cdot 5$, while the relations corresponding to (96) show that $a_n = (-\cdot 5)^n$. Considering the general formula for the regular process,

$$\{\zeta(t)\} = \{\eta(t)\} + \cdot 5\,\{\eta(t-1)\},$$

it follows that

$$\{\eta(t)\} = \{\zeta(t)\} - \cdot 5\,\{\zeta(t-1)\} + (\cdot 5)^2\{\zeta(t-2)\} - (\cdot 5)^3\{\zeta(t-3)\} + \cdots,$$

in full agreement with (200). Observing that $K^2 = 1\cdot 25$, and that $(K^{(1)})^2 = 5$, formula (258) gives for the remaining process in the group $(\zeta^{(i)})$

$$\{\zeta^{(1)}(t)\} = \cdot 5\,\{\eta(t)\} + \{\eta(t-1)\}. \tag{264}$$

According to the general analysis, the residual $\{\eta^{(1)}(t)\}$ secured by theorem 6 is obtained by replacing ζ by $\zeta^{(1)}$ in (200). As is readily verified, the corresponding representation of type (260) reads

$$\eta^{(1)}(t) = \tfrac{1}{2}\,\eta(t) + \tfrac{3}{4}\,\eta(t-1) - \tfrac{3}{8}\,\eta(t-2) + \tfrac{3}{16}\,\eta(t-3) - \cdots$$

We obtain, in analogy with (264), and in harmony with the general formula (261)

$$\{\zeta^{(1)}(t)\}=\{\eta^{(1)}(t)\}+\cdot 5\,\{\eta^{(1)}(t-1)\}=\cdot 5\,\{\eta(t)\}+\{\eta(t-1)\}.$$

It can be easily verified that $D^2(\eta^{(1)}(t))=D^2(\eta(t))$, and that the process $\{\eta^{(1)}(t)\}$ is non-autocorrelated, i. e. that $r_k(\eta^{(1)})=0$ for $k \neq 0$.

3) Next, let $r_1=\frac{1}{6}$; $r_2=-\frac{1}{3}$; $r_k=0$ for $k>2$.

The relation (256) reads in this case

$$\tfrac{2}{3}(x^2+\cdot 5x-\cdot 5)(-\cdot 5x^2+\cdot 5x+1)=-\tfrac{1}{3}x^4+\tfrac{1}{6}x^3+x^2+\tfrac{1}{6}x-\tfrac{1}{3}.$$

Paying regard to the identity $x^2+\cdot 5x-\cdot 5=(x-\cdot 5)(x+1)$, a short calculation will show that the group $(b^{(i)})$ consists of two sequences, viz. $(1, \cdot 5, -\cdot 5)$, which is the regular one, and $(1, -1, -2)$. The general formula for a corresponding regular process reads

$$\{\zeta(t)\}=\{\eta(t)\}+\cdot 5\,\{\eta(t-1)\}-\cdot 5\,\{\eta(t-2)\}.$$

We find without difficulty

$$a_k=\tfrac{1}{3}(\tfrac{1}{2})^k+\tfrac{2}{3}(-1)^k.$$

The non-regular process is given by

$$\begin{aligned}\{\zeta^{(1)}(t)\}&=\cdot 5\,\{\eta(t)\}-\cdot 5\,\{\eta(t-1)\}-\{\eta(t-2)\}=\\&=\{\eta^{(1)}(t)\}+\cdot 5\,\{\eta^{(1)}(t-1)\}-\cdot 5\,\{\eta^{(1)}(t-2)\}.\end{aligned}$$

Formula (260) gives for the non-autocorrelated residual

$$\eta^{(1)}(t)=\tfrac{1}{2}\eta(t)-\tfrac{3}{4}\eta(t-1)-\tfrac{3}{8}\eta(t-2)-\tfrac{3}{16}\eta(t-3)-\cdots$$

4) A few remarks in connexion with the illustrations concluding Chapter II will be sufficient to show how the general formulae given in the present chapter will work in the case of a normal process of moving averages. The developments below cover both processes of linear autoregression and processes of moving averages.

The matrix of the infinite quadratic form appearing in the characteristic function of a non-autocorrelated normal process $\{\eta\}$ with dispersion 1 is nothing else than the unit matrix. Now, keeping in mind the substitution procedure indicated on p. 90, the relations (205) will verify the product formula already given on p. 91,

$$\begin{pmatrix}1 & b_1 & b_2 & \cdot\,\cdot\\ 0 & 1 & b_1 & \cdot\,\cdot\\ 0 & 0 & 1 & \cdot\,\cdot\\ \cdot & \cdot & \cdot & \cdot\,\cdot\end{pmatrix}\cdot\begin{pmatrix}\varkappa^2 & 0 & 0 & \cdot\,\cdot\\ 0 & \varkappa^2 & 0 & \cdot\,\cdot\\ 0 & 0 & \varkappa^2 & \cdot\,\cdot\\ \cdot & \cdot & \cdot & \cdot\,\cdot\end{pmatrix}\cdot\begin{pmatrix}1 & 0 & 0 & \cdot\,\cdot\\ b_1 & 1 & 0 & \cdot\,\cdot\\ b_2 & b_1 & 1 & \cdot\,\cdot\\ \cdot & \cdot & \cdot & \cdot\,\cdot\end{pmatrix}=\begin{pmatrix}1 & r_1 & r_2 & \cdot\,\cdot\\ r_1 & 1 & r_1 & \cdot\,\cdot\\ r_2 & r_1 & 1 & \cdot\,\cdot\\ \cdot & \cdot & \cdot & \cdot\,\cdot\end{pmatrix},$$

which yields the matrix belonging to a normal process of finite or infinite moving averages. The previous illustrations exemplify the fact that there are, in general, several sequences (b) giving rise to the same matrix in the right member.

The general inversion formula (200) implies

$$(265)\qquad \begin{pmatrix} 1 & a_1 & a_2 & \cdot\cdot \\ 0 & 1 & a_1 & \cdot\cdot \\ 0 & 0 & 1 & \cdot\cdot \\ \cdot & \cdot & \cdot & \cdot\cdot \end{pmatrix} \cdot \begin{pmatrix} 1 & r_1 & r_2 & \cdot\cdot \\ r_1 & 1 & r_1 & \cdot\cdot \\ r_2 & r_1 & 1 & \cdot\cdot \\ \cdot & \cdot & \cdot & \cdot\cdot \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & \cdot\cdot \\ a_1 & 1 & 0 & \cdot\cdot \\ a_2 & a_1 & 1 & \cdot\cdot \\ \cdot & \cdot & \cdot & \cdot\cdot \end{pmatrix} = \begin{pmatrix} \varkappa^2 & 0 & 0 & \cdot\cdot \\ 0 & \varkappa^2 & 0 & \cdot\cdot \\ 0 & 0 & \varkappa^2 & \cdot\cdot \\ \cdot & \cdot & \cdot & \cdot\cdot \end{pmatrix}.$$

The verification is immediate. In fact, paying regard to (208) and (210) we obtain in the first place

$$\begin{pmatrix} 1 & r_1 & r_2 & \cdot\cdot \\ r_1 & 1 & r_1 & \cdot\cdot \\ r_2 & r_1 & 1 & \cdot\cdot \\ \cdot & \cdot & \cdot & \cdot\cdot \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & \cdot\cdot \\ a_1 & 1 & 0 & \cdot\cdot \\ a_2 & a_1 & 1 & \cdot\cdot \\ \cdot & \cdot & \cdot & \cdot\cdot \end{pmatrix} = \begin{pmatrix} \varkappa^2, & \varkappa^2 \cdot b_1, & \varkappa^2 \cdot b_2, & \cdot\cdot \\ 0\,, & \varkappa^2\,, & \varkappa^2 \cdot b_1, & \cdot\cdot \\ 0\,, & 0\,, & \varkappa^2\,, & \cdot\cdot \\ \cdot & \cdot & \cdot & \cdot\cdot \end{pmatrix}.$$

Premultiplying this relation with the first factor of (265), and keeping in mind the relations (201), we arrive at (265).

5) Let p_i be the multiplicity of a real root or of a conjugate complex root-pair of (251). Then the number of averages in the same group as (249) is $\prod_i (p_i + 1)$, the index i running over those roots and root-pairs which are $\neq 1$ in modulus.

CHAPTER IV.

On the application of some stationary schemes.

27. Preliminary remarks. Disposition.

In applied time series analysis, the chief problem is to find a hypothetical scheme which from a theoretical viewpoint is appropriate to the phenomenon considered, and gives a satisfactory fit to the observational data. Another desideratum is, of course, that the hypothesis be as simple as possible.

As shown in detail in sections 15 and 16, the general stationary process embraces all of the hypotheses about time series surveyed in Chapter I. The wide scope of the stationary process is due to the fact that the restrictions are reduced to a minimum: Besides the indispensable postulate that the probability laws must not contradict themselves (see (53)—(54)), the only further assumption is that time itself will not influence these probability laws (see (55)). In other words, time is thought of as a passive medium; roughly speaking, this means that any prognosis based upon the past development will depend only on this same development — i.e. supposing that the same development had taken place with a constant lag, the corresponding forecast would not differ, apart from the displacement in time.

It stands to reason that the assumption of stationarity is legitimate in the most varied fields of scientific research. Then if we restrict the analysis to equidistant time points, we have at our disposal all of the schemes falling under the discrete stationary process. Regarding the simplification of treating time as a discrete parameter, it needs no comment that this device is appropriate in most if not all practical applications (cf. p. 70 f).

From the viewpoint of the theory of probabilities, the purely random process is the simplest type case of a stationary scheme. In sections 15 and 16, certain other type cases were constructed

on the basis of the purely random process, and the scheme of hidden periodicities was interpreted as a stationary process. The special theoretical models mentioned are rather simple, in as much as an adequate description of their structure is possible by means of linear methods, such as the periodogram analysis and the analysis of the graph of autocorrelation coefficients.

The present chapter is reserved for some applications of the above-mentioned simple schemes, in particular the schemes of linear regression. It will be seen that the results obtained are rather promising. We advance also that certain points in the applications will give rise to theoretical discussions — the analysis in the previous chapters was chiefly concerned with the structural properties of the hypothetical models, while questions bearing upon their application were touched upon only incidentally.

Of course, in applying different hypothetical schemes to observational data, different methods are required.

In the search for a hypothetical model suitable to a stationary phenomenon, the construction of an empirical periodogram is a classical method of fundamental importance. A careful periodogram construction is a safe method for discovering hidden periodicities if such are really present. On the other hand, approximate methods often involve definite dangers. For instance, the BRUNS-OPPENHEIM method (see section 5) fails as often as the periodic elements are covered by a random component. The bias in question, which was already observed by J. I. CRAIG (1916), is of interest also in view of other methods of analysis. For this reason, the »CRAIG effect» will be examined in some detail. This is done in the next section.

As shown in detail in Chapter III, periodogram analysis is an inadequate method of research in the cases of linear autoregression and of moving averages. The present study being focussed on these schemes, we have to look for appropriate substitutes for the periodogram method. In the memoir where G. U. YULE (1927) introduces the scheme of linear autoregression, an empirical parallel to the autoregression analysis as developed in section 19 forms the leading method of research. This method yields a first substitute for the periodogram construction. Next, as emphasized by SIR G. WALKER (1931), the autocorrelation coefficients behave quite differently in the schemes of hidden periodicities and linear autoregression. Hence, the graph of the serial coefficients will yield important information

about the nature of the phenomenon considered. The analysis in section 26 shows that the scheme of moving averages, too, presents a characteristic graph of autocorrelation coefficients, a circumstance which augments the importance of the method proposed by WALKER. For the sake of brevity in writing, the graphs of serial and autocorrelation coefficients will be termed *correlograms* (*empirical* and *hypothetical* respectively).

In the following, the methods proposed by G. U. YULE (1927) and SIR G. WALKER (1931) will be carried further on the basis of the theoretical investigations in the previous chapters, and used in the applications to empirical data. A critical survey of the original methods of YULE and WALKER will be given in section 29. In section 30 follow a summary and a critical examination of the modified methods.

The two sections concluding the present study are reserved for applications of the scheme of moving averages and the scheme of linear autoregression. Both these schemes of linear regression may be attached directly to familiar lines of time series analysis. Returning to this point later, especial attention will be drawn to the different type of forecast yielded by these schemes as compared with the scheme of hidden periodicities. While the forecasts obtained in the latter scheme cover an infinite future, the former schemes will yield efficient forecasts only over a limited period of time. However, this limitation is outweighed by the greater efficiency in the short time forecasts yielded by the schemes of linear regression.

Finally, in discussing the applications, certain generalizations of the schemes of linear regression will be touched upon.

28. On the CRAIG effect.

In section 5 we have surveyed a few methods for separating the individual components in a sum of harmonics (17). The classical method being the construction of a periodogram, the short cut indicated by S. OPPENHEIM (1909) (see p. 18) is based on the difference relations satisfied by the function (17).

The scheme of hidden periodicities consists of a composed harmonic on which a random component is additively superposed.

Even in this case a periodogram will point out the frequencies of the individual harmonics. But as emphasized by J. I. Craig (1916), the Oppenheim method will be biassed by the random component. Proceeding to an examination of this bias, which will be termed the »Craig effect», it will be sufficient for our purpose to consider a scheme of simple structure. In doing this, we shall regard the composed harmonic as a sample series of a singular process. This device, which will not affect the proof, is used merely in order to illustrate the connexion between the scheme of hidden periodicities as defined in section 8 and the process of hidden periodicities as defined in section 15 Ω.

Let $\{\psi(t)\}$ stand for a singular process satisfying the relation

$$\Delta^2 \psi(t-1) + k^2 \cdot \psi(t) = 0, \; 0 < k \leq 2.$$

Let $\{\eta(t)\}$ be purely random, and independent of $\{\psi(t)\}$. Further let $\{\eta(t)\}$ have a finite dispersion and a vanishing mean, and let a process of hidden periodicities be defined by

$$\{\xi(t)\} = \{\psi(t)\} + \{\eta(t)\}. \tag{266}$$

Denoting by $\psi_i(t, t-1, \ldots, t-n) = [\psi_i(t), \psi_i(t-1), \ldots, \psi_i(t-n)]$ a sample series section of the process $\{\psi(t)\}$, we have

$$\Delta^2 \psi_i(t-s-1) + k^2 \psi_i(t-s) = E[\Delta^2 \psi(t-1) + k^2 \psi(t)] = 0.$$

Now, let k^2 be unknown. The value delivered by the Oppenheim method is that minimizing the expression

$$\frac{1}{n-1} \sum_{s=1}^{n-1} [\Delta^2 \psi_i(t-s-1) + k^2 \psi_i(t-s)]^2.$$

A short reduction leads to the following value which is obviously unbiassed,

$$k^2 = -\sum_{s=1}^{n-1} \psi_i(t-s) \cdot \Delta^2 \psi_i(t-s-1) \Big/ \sum_{s=1}^{n-1} \psi_i^2(t-s) = \tag{267}$$

$$= -E[\psi(t) \cdot \Delta^2 \psi(t-1)] \,/\, E[\psi^2(t)].$$

According to (31), the frequency λ of the harmonic $\psi_i(t)$ is given by

$$\lambda = 2 \text{ arc sin } k/2. \tag{268}$$

Let it next be assumed that a sample series $\xi_i(t, t-1, .., t-n)$ is given, and that we know neither k^2 nor the sample series ψ_i and η_i. The assumption concerning an additive random element corresponds to the actual situation when seeking for a period in an empirical time series. Applying now the OPPENHEIM method, we must minimize

$$\text{(269)} \qquad \frac{1}{n-1} \cdot \sum_{s=1}^{n-1} [\Delta^2 \xi_i(t-s-1) + \bar{k}^2 \cdot \xi_i(t-s)]^2,$$

which gives (cf. (267))

$$\text{(270)} \qquad \bar{k}^2 = -\sum_{s=1}^{n-1} \xi_i(t-s) \cdot \Delta^2 \xi_i(t-s-1) \Big/ \sum_{s=1}^{n-1} \xi_i^2(t-s).$$

This expression depends on the actual path of the sample series $\xi_i(t, t-1, .., t-n)$, but a sufficient approximation is delivered by the $\bar{k}^2$-value minimizing the expression

$$\text{(271)} \qquad E_i[\Delta^2 \psi(t-1) + \bar{k}^2 \psi(t)]^2 + E[\Delta^2 \eta(t-1) + \bar{k}^2 \eta(t)]^2,$$

where we have written $E_i[f[\psi]]$ for $\lim\limits_{n\to\infty} \frac{1}{n-1} \cdot \sum\limits_{s=1}^{n-1} f[\psi_i(t-s)]$. A short calculation shows that this approximation reads

$$\text{(272)} \qquad \bar{k}^2 \sim [-E_i \psi(t) \cdot \Delta^2 \psi(t-1) + 2\, E \eta^2(t)] \,/\, [E_i \psi^2(t) + E \eta^2(t)].$$

The difference between on one hand (270) and (272) and on the other hand the unbiassed formula (267) gives rise to the CRAIG effect. Formula (272) shows that this effect will depend on the chance-determined amplitude of the harmonic constituting $\psi_i(t, t-1, .., t-n)$.

It is seen that the $\bar{k}^2$-value given by (272) equals k^2 when, and only when, $k^2 = 2$, i. e. when $\psi(t)$ has the period 4 time units. A brief reflection shows further that the OPPENHEIM method will under-estimate the period if this is above 4 time units, while the reverse will be true if the period is below 4 time units. Moreover, the larger the variance of the random component η, the larger is the CRAIG effect.

Illustrations. Two simple examples of the CRAIG effect will be given on the basis of the model series $(\bar{\Omega}_t^{(1)})$ and $(\bar{\Omega}_t^{(2)})$ presented in table 4. Applying the relations (270), (268) and (18), the following results were obtained, the sums running from $t = 2$ to $t = 999$.

	Model series ($\overline{\Omega}_t^{(1)}$)	Model series ($\overline{\Omega}_t^{(2)}$)
$\Sigma\, \overline{\Omega}_t \cdot \Delta^2\, \overline{\Omega}_{t-1}$	-4041	-5568
$\Sigma\, [\overline{\Omega}_t]^2$	1112	1711
$\overline{k}^2$	3·6340	3·2542
$\overline{\lambda}$	144·8°	128·8°
$\overline{p}$	2·486	2·795

The correct value of the period being 2 units in both the time series, the CRAIG effect is seen to be substantial. The values $\overline{k}^2$ obtained are in good agreement with the approximate formula (272). In fact, observing that in each of the two cases $E_i[\psi(t) \cdot \Delta^2 \psi(t-1)] = -4$, and that the variances of the random components are respectively ·2 and ·6, we get in the first case $\overline{k}^2 \sim 4 \cdot 4/1 \cdot 2 = 3 \cdot 667$, and in the second $\overline{k}^2 \sim 5 \cdot 2/1 \cdot 6 = 3 \cdot 25$. In full agreement with the remarks attached to formula (272), the period is over-estimated, and the CRAIG effect is larger in the second model series than in the first.

In the above analysis of the scheme (266) we have examined the OPPENHEIM method for determining the parameter k in an approach of type $\Delta^2 \xi(t-1) \sim k \cdot \xi(t)$ or $\xi(t) \sim (2-k) \cdot \xi(t-1) - \xi(t-2)$. We shall next apply the same method in starting from a more general approach, viz.

(273) $$\xi(t) \sim a_1 \cdot \xi(t-1) + a_2 \cdot \xi(t-2).$$

Letting as before $\xi_i(t, t-1, .., t-n)$ represent a sample series section connected with the scheme (266) of hidden periodicities, we must in the present case minimize (cf. (269))

(274) $$\frac{1}{n-1} \sum_{s=0}^{n-2} [\xi_i(t-s) - \bar{a}_1 \cdot \xi_i(t-s-1) - \bar{a}_2 \cdot \xi_i(t-s-2)]^2$$

in respect of $\bar{a}_1$ and $\bar{a}_2$. Now, using an approximation of the same type as in (272), and paying regard to (46), we get

(275) $$r_1 - \bar{a}_1 - \bar{a}_2 \cdot r_1 \sim 0, \quad r_2 - \bar{a}_1 \cdot r_1 - \bar{a}_2 \sim 0,$$

having written $r_k = \cos \lambda k / (1 + d_i^2)$. Here λ is the frequency of a sample series connected with the singular process $\{\psi(t)\}$, and $d_i^2 = D^2(\eta) / D_i^2(\psi)$, where $D_i(\psi)$ is the dispersion in the sample series $\psi_i(t, t-1, \ldots)$. Solving (275), we obtain

$$(276)\qquad \begin{aligned} \bar{a}_1 &\sim 2 \cdot \frac{\sin^2 \lambda + d_i^2}{\sin^2 \lambda + 4\, d_i^2 + 4\, d_i^4} \cdot \cos \lambda; \\ \bar{a}_2 &\sim -\frac{\sin^2 \lambda - 2\, d_i^2 \cdot \cos 2\lambda}{\sin^2 \lambda + 4\, d_i^2 + 4\, d_i^4} \geq -1. \end{aligned}$$

The minimizing of (274) is obviously analogous to that phase of the autoregression analysis described in section 19, where $\xi(t)$ is linearly approximated by $\xi(t-1)$ and $\xi(t-2)$. It is in view of this analogy that the above formulae are of interest. Now, if $D(\eta) = 0$, we obtain from (276) the coefficients $\bar{a}_1 = 2 \cos \lambda$ and $\bar{a}_2 = -1$ appearing in the identity (cf. (35))

$$\xi(t) = \psi(t) = 2 \cos \lambda \cdot \psi(t-1) - \psi(t-2).$$

It should further be observed that we have in this case $\cos \lambda = \bar{a}_1 / 2\sqrt{-\bar{a}_2}$. On the other hand, if $D(\eta) \neq 0$, the latter relation will be disturbed by a Craig effect. In fact, we get from (276)

$$(277)\qquad \cos \bar{\lambda} = \frac{(\sin^2 \lambda + d_i^2) \cdot \cos \lambda}{\sqrt{(\sin^2 \lambda - 2\, d_i^2 \cdot \cos 2\lambda)(\sin^2 \lambda + 4\, d_i^2 + 4\, d_i^4)}}.$$

Approximating the period p of $\psi(t)$ by means of the biased frequency $\bar{\lambda}$ given by (277), the Craig effect is seen to be particularly large if $\sin \lambda$ is small. It would serve no purpose to discuss the sign of the deviation or to enter into details on a singular process $\{\psi(t)\}$ of general structure.

Illustration. Considering the model series $(\bar{\Omega}_t^{(1)})$ dealt with in the previous illustration, we have found by minimizing (274) $-910 = \bar{a}_1 \cdot 1114 - \bar{a}_2 \cdot 911$; $930 = -\bar{a}_1 \cdot 911 + \bar{a}_2 \cdot 1114$. This system gives $\bar{a}_1 = -\cdot 4051$, $\bar{a}_1 = \cdot 5036$. Since $a_1 = -2$, and $a_2 = -1$ the method examined is completely misleading. Formula (276) explains the failure — in the present case $\sin \lambda = 0$, and the resulting $\bar{a}$-values will show no tendency whatever to approximate the a-values sought for. As is readily verified, (276) gives $-\bar{a}_1 \sim \bar{a}_2 \sim \cdot 4167$.

Summing up the above analysis, we conclude that the Oppenheim method yields no adequate substitute for a periodogram construction in case the periodicities are covered by a random element, and it does not seem worth the trouble to derive modifications neutralizing the bias involved.

29. On earlier applications of the scheme of linear autoregression.

As far as I am aware, there are only two earlier investigations which present direct applications of the schemes of linear regression to empirical data, viz. those already referred to by G. U. YULE (1927) and SIR G. WALKER (1931). These memoirs are concerned with sunspots and air pressure respectively, and in both cases it is the scheme of autoregression that is applied. An examination of the main lines of these investigations in the light of the previous analysis will be given in the present section. For the sake of completeness, we shall also touch upon a passage in the already mentioned study on expectance theory by K. STUMPFF (1936), where an empirical correlogram is dealt with by use of a method related to that proposed by SIR G. WALKER.

The memoir of G. U. YULE (1927) starts with a discussion of a model series, say $\bar{\zeta}_t$, constructed on the basis of a relation of type

$$\bar{\zeta}_t - k\,\bar{\zeta}_{t-1} + \bar{\zeta}_{t-2} = \bar{\eta}_t, \quad -2 < k < 2, \tag{278}$$

The purely random series $\bar{\eta}_t$ is obtained by dice-throwing. The constant k is chosen in the interval $(-2, 2)$, which implies that the roots of the characteristic equation of (278) are complex, and of modulus unity. Thus it follows from the general analysis in section 22 that a process $\{\zeta(t)\}$ corresponding to (278) is non-stationary. However, the evolutive tendency is rather weak, and the 300 elements constituting YULE's model series actually present fluctuations of a stationary appearance.

The parameter k in the model series $\bar{\zeta}_t$ being chosen so as to correspond to a period of 10 time units, YULE lays stress upon the structural resemblance between his model series and the yearly index of sunspots. Pursuing this suggestion in the later sections of his memoir, he gives two different methods for a refined analysis of the structure of the index. YULE works on the A. WOLFER index 1751—1923.

In his first approach, YULE starts from the hypothesis that the sunspot index, say $\bar{\zeta}_t$, satisfies a relation of type (278). In order to determine k, he minimizes the sum of the squared »disturbances» $\bar{\eta}_t$. Interpreting the relation (278) as ruling the movement of a

pendulum subjected to random shocks, YULE derives the intrinsic period of the pendulum which would correspond to the k-value obtained. The period thus derived being too short, viz. $p = 10{\cdot}08$ years, he finds that the hypothesis (278) gives a better value, $p = 11{\cdot}03$ years, when applied after graduating the sunspot numbers.

Having thus assumed a linear regression (278) between $\bar{\zeta}_t + \bar{\zeta}_{t-2}$ and $\bar{\zeta}_{t-1}$, YULE applies a graphic test of this approach. Referring to the graphs given on p. 277, he says on the same page: »On the whole, however, divergence from linearity does not look as if it would be a serious trouble».

In his second approach, YULE starts from the relation

$$\bar{\zeta}_t + a_1 \bar{\zeta}_{t-1} + a_2 \bar{\zeta}_{t-2} = \bar{\eta}_t. \tag{279}$$

Proceeding as in the case (278), he determines the parameters (a) by minimizing $\Sigma\, \bar{\eta}_t^2$, and interprets the results by the use of the analogy with a pendulum. The values found for a_1 and a_2 correspond to a damped intrinsic oscillation. The ungraduated index gives the period $p = 10{\cdot}600$ years, while the graduated index as before gives a better value, $p = 11{\cdot}164$ years.

The graduated index gives rise to a smaller variance in the disturbances $\bar{\eta}_t$ than the ungraduated index. Dividing the variance of $\bar{\eta}_t$ by the variance of the sunspot index under analysis, the approach (278) gives $\cdot 243$ for the ungraduated index, and $\cdot 115$ for the graduated one. The corresponding values in the approach (279) are $\cdot 198$ and $\cdot 102$ respectively.

In applying generalized hypotheses of type (278) and (279), YULE finds that the introduction of more parameters does not bring on a marked decrease in the variance of the disturbances. In other words, the experiments »fail to suggest the presence of any period other than the fundamental, a conclusion entirely in accord with the work of LARMOR and YAMAGA» (p. 295).

In a summary, YULE suggests that the sunspot numbers »should be regarded as analogous to the data that would be given by observations of a disturbed periodic movement, such as that of a pendulum subjected to successive small random impulses» (p. 294). Let us discuss this hypothesis in the light of the previous analysis.

As already pointed out, the approach (278) does not correspond to a stationary process, for if the series $\bar{\eta}_t$ were purely random, the secondary model series $\bar{\zeta}_t$ would present oscillations increasing in amplitude with time, i. e. be evolutive. In view of this observation,

it is not surprising that the disturbances $\bar{\eta}_t$, as calculated from (278) on the basis of the k-value previously determined, form a series of non-random character. Quoting YULE, the series $\bar{\eta}_t$ shows »a tendency for positive disturbances during the approach to the maximum of the sunspot numbers, negative during the approach to minimum» (p. 294 f.).

We have already seen that the approach (279) gives rise to a somewhat smaller variance in the disturbances $\bar{\eta}_t$. However, having introduced a second parameter, the slight reduction in the variance of $\bar{\eta}_t$ does not give a sufficient reason for preferring (279) over (278). On the other hand, (279) corresponds to a proper stationary process, a circumstance speaking in favour of this approach.

In itself, the outcome of significant constants a_1 and a_2 does not imply that the approach (279) is adequate. Without further evidence, we cannot even conclude that (279) is better than the hypothesis of a strictly periodic component in the sunspot index. In fact, our analysis in the previous section has shown that an autoregression analysis will give rise to non-vanishing coefficients (a) also in the case of hidden periodicities. It is interesting to notice that even the effect of the graduation — the increase in the period — might be explained by assuming the index to be ruled by a scheme of hidden periodicities (cf. p. 137). However, a sufficient reason for rejecting the latter hypothesis is that the deduction of a strictly periodic component actually leaves an »error» with a dispersion substantially above that of the disturbances obtained from (279). In fact, the assumption of one harmonic component in the sunspot index would explain at most 28 % of the variance in the index (see e. g. K. STUMPFF (1937), p. 126). On the other hand, we have seen that the approach (279), which contains only two parameters, is able to explain at least 80 % of the variance in question. In this connexion it is rather interesting to notice that the periodogram of the sunspot index given by STUMPFF (l. c.) bears a certain resemblance to our fig. 4 (p. 116), and thus agrees with the hypothesis of linear autoregression. (Cf. also the remarks attached to (131)).

The values found by YULE for the parameters in (279) are $a_1 = -1{\cdot}34254$, $a_2 = {\cdot}65504$ for the ungraduated, and $a_1 = -1{\cdot}51527$, $a_2 = {\cdot}80245$ for the graduated sunspot index. Using the analogy of a swinging pendulum, these values correspond to a rather heavy damping — in the case $a_2 = {\cdot}80245$, the amplitude of a swing

would be reduced to 29 % in the duration of one period (see G. U. Yule (1927), p. 282). With this heavy damping, a purely random series of impulses would not be likely to produce such large amplitudes as in the sunspot fluctuations. In full agreement with this argument, which in section 32 will be developed as a test of the scheme of linear autoregression, the disturbances $\bar{\eta}_t$ calculated from (279) on the basis of the sunspot index present variations of a non-random type. We conclude that the situation is not quite covered by Yule's statement that »the disturbances do occur just in the kind of way that would be necessary to maintain a damped vibration» (p. 286).

Summing up the above discussion, we have seen that in replacing the approach of strict periodicity by a hypothesis containing an acting random element, G. U. Yule (1927) obtains a substantially better fit to the sunspot data. In the terminology of the present study, the approach (279) as applied to the ungraduated index corresponds to a scheme of linear autoregression. On the other hand, as applied to the graduated index, it is obvious that the model (279) approximately corresponds to the assumption that the ungraduated index is ruled by a scheme consisting in a purely random process independent of and superposed on a process of linear autoregression.

As mentioned by Yule, the disturbances $\bar{\eta}_t$ calculated from (279) present a certain systematic variation. This non-random behaviour, which seems to be conditioned by the small value found for a_2 in (279), remains unexplained by the hypothesis of linear autoregression. In view of this circumstance, it seems to me as if the sunspot index calls for further investigation. Perhaps the methods developed in section 32 would yield a scheme fitting the data better. But it is also possible that more satisfactory results would be obtained in an approach involving a non-linear function of the index. Since to pursue these suggestions falls outside the program of this survey, we shall end our discussion of the Yule memoir.

Sir G. Walker (1931) follows up the Yule approach (279), and studies an autoregression relation of type

$$\bar{\zeta}_t + a_1 \bar{\zeta}_{t-1} + \cdots + a_h \bar{\zeta}_{t-h} = \bar{\eta}_t. \tag{280}$$

As mentioned in section 24, Walker finds that the autocorrelation coefficients corresponding to (280) satisfy the relations (cf. p. 104)

$$(281) \qquad r_k + a_1 r_{k-1} + \cdots + a_h r_{k-h} = 0, \; k > h.$$

Assuming that the roots of the characteristic equation (34) are different, he further gives r_k as the general solution of (32).

We are now in a position to examine WALKER's methods for applying (280) to empirical data. The basic idea is simply to compute the serial coefficients $\bar{r}_k$, and to determine the constants a_n so that the relations (281) will be approximately satisfied when replacing r_k by $\bar{r}_k$.

SIR G. WALKER works on air pressure data from *Port Darwin* 1880—1925, taking the quarter of a year for time unit. The graph of serial coefficients — in our terminology the correlogram — ranges from $k = 1$ to $k = 147$ quarters (p. 531). The graph shows a rapid decrease from $\bar{r}_1 = \cdot 76$ to $\bar{r}_6 \cong 0$. For larger k-values, the correlogram presents fluctuations with rather small amplitudes. In fact, up to $k = 100$ all the coefficients $\bar{r}_k$ are less than $\cdot 3$ in modulus.

SIR G. WALKER finds that in the interval $0 \leq k \leq 40$ a fairly good approximation to the correlogram is yielded by the function

$$(282) \qquad r_k = \cdot 19 (\cdot 96)^k \cos \pi k / 6 + \cdot 15 (\cdot 98)^k + \cdot 66 (\cdot 71)^k.$$

This function, which is seen to involve a damped harmonic with period $p = 12$ quarters, satisfies the difference equation

$$r_k - 3 \cdot 35\, r_{k-1} + 4 \cdot 43\, r_{k-2} - 2 \cdot 71\, r_{k-3} + \cdot 64\, r_{k-4} = 0.$$

Concluding inversely from (281) on (280), WALKER finally arrives at the representation

$$(283) \qquad \bar{\zeta}_t - 3 \cdot 35\, \bar{\zeta}_{t-1} + 4 \cdot 43\, \bar{\zeta}_{t-2} - 2 \cdot 71\, \bar{\zeta}_{t-3} + \cdot 64\, \bar{\zeta}_{t-4} = \bar{\eta}_t.$$

Proceeding to an examination of WALKER's methods, we observe in the first place that an unconditioned conclusion from (281) on (280) is not permitted. In fact, the autocorrelation coefficients corresponding to the relation (280) satisfy not only the relations (281) but also (222—223), the latter relations not having been observed by WALKER. It is seen that the coefficients $r_1, r_2, \ldots r_{h-1}$ will be uniquely determined by the system (222) in terms of the coefficients a_n. In other words, the coefficients in (280) determine not only the periods and the damping factors of the individual components in the expression (33), but also the coefficients of the

components. Thus, without anything further we cannot be sure that the autocorrelation coefficients corresponding to the approach (283) will be given by the graduating expression (282). On the contrary, the system (222) corresponding to (283) actually gives rise to autocorrelation coefficients which are substantially different from (282). Having read off the values $r_1 = \cdot 75$, $r_2 = \cdot 55$, $r_3 = \cdot 35$ from the graph of (282) given by WALKER (p. 528), I have found from (222) the values $r_1 = \cdot 93$, $r_2 = \cdot 72$, $r_3 = \cdot 43$ for the autocorrelation coefficients belonging to the approach (283).

In view of the oversight pointed out, it is not surprising that the relation (283) gives rise to a larger dispersion in the disturbances $\bar{\eta}_t$ than in the air pressure data $\bar{\zeta}_t$. As a matter of fact, I have found $\bar{D}(\bar{\eta}) = 2 \cdot 4 \ \bar{D}(\bar{\zeta})$, while the result $\bar{D}(\bar{\eta}) = \cdot 28 \ \bar{D}(\bar{\zeta})$ obtained by WALKER (p. 530) is based on an incorrect use of relation (317) (see also our foot-note remark on p. 112). We conclude that if the approach (283) is to be applied to the air pressure data, the coefficients must be modified. Having stated this, it is rather interesting that the simple approach

$$(284) \qquad \bar{\zeta}_t - \cdot 73 \, \bar{\zeta}_{t-1} = \bar{\eta}_t$$

gives a fairly good fit to the first few serial coefficients. In fact, according to (238) the approach (284) gives $r_1 = \cdot 73$, $r_2 = \cdot 53$, $r_3 = \cdot 39$, $r_4 = \cdot 28$, while the air pressure serial coefficients given by WALKER (p. 528) read $\bar{r}_1 = \cdot 76$, $\bar{r}_2 = \cdot 56$, $\bar{r}_3 = \cdot 36$, $\bar{r}_4 = \cdot 18$. A short calculation shows further that (284) gives $D(\eta) = \cdot 68 \, D(\zeta)$.

Let us in conclusion attach a few remarks to the empirical correlogram presented by SIR G. WALKER (p. 531). As already mentioned, the serial coefficients show rather small deviations from zero in the interval $3 < k < 40$. On the other hand, the increase in amplitude for certain k-values > 40 might be due to the successive reduction in the number of correlates. Perhaps this argument is sufficient to explain also why the fluctuations are somewhat larger in that alternative variant of a correlogram given by WALKER, where all serial coefficients are based on 77 pairs of correlates. As the fluctuations, furthermore, seem rather irregular and aperiodic — at least to my eye — it is doubtful whether it would be possible to improve sensibly the approach (284) by taking into account more distant elements $\bar{\zeta}_{t-2}$, $\bar{\zeta}_{t-3}$, etc. In this connexion it is rather interesting to notice that according to the general analysis there exists no process of linear autoregression having (282) for auto-

correlation coefficients. Another reason for resting satisfied with the simple approach (284) is that the ordinates in the periodogram presented by SIR G. WALKER (p. 526) are all lying on about the same level — this periodogram does not like that of the sunspots suggest a scheme of linear autoregression with a tendency to periodicity (cf. p. 142). At any rate, a more detailed analysis of the air pressure data is beyond the scope of the present survey.

Using formula (127), K. STUMPFF (1936) develops a periodogram theory which generalizes the classical SCHUSTER theory. In applying his theory, STUMPFF works on air pressure data (*Potsdam* $^3/_7$ 1925—$^{27}/_6$ 1926, equidistance 1 hour), and replaces the coefficients r_k in (127) by a corresponding set of graduated serial coefficients. Claiming that the graduated values belong to a scheme of linear autoregression of type

$$(285) \qquad \zeta(t) - 2p \cdot \zeta(t-1) + p^2 \cdot \zeta(t-2) = \eta(t),$$

STUMPFF makes a mistake similar to WALKER's pointed out above. Proceeding as in the developments on p. 112, I have found the formula $r_k = [1 + k \cdot (1 - p^2)/(1 + p^2)] \cdot p^k$ for the autocorrelation coefficients belonging to the scheme (285), while K. STUMPFF (p. 53) gives $r_k = (1 - k \cdot \log p) \cdot p^k$.

30. Preliminary survey of methods.

In this section we shall give a brief summary of the methods used in the later applications and a few critical remarks on the scope of these methods.

As pointed out in earlier sections, a careful analysis of the structural properties of a time series requires statistical data covering a rather long period. On the other hand, the series must not change its general character in the course of the interval of observation, for then a stationary scheme would be inadequate. For instance, if a trend is present in the material, it should be removed before starting the analysis (cf. p. 1).

Considering a scheme (39) of hidden periodicities, and disregarding the value $r_0 = 1$, the correlogram r_k consists of superposed harmonics such that the periods of the individual components equal those in the time series considered (cf. (46)). In the two type cases of linear

regression, on the other hand, the correlogram has the horizontal axis for asymptote. In fact, in the scheme of linear autoregression the correlogram r_k is a function (33) such that the roots of the characteristic equation (34) are of modulus less than unity, and in the scheme of moving averages all the autocorrelation coefficients are zero beyond a certain k-value (cf. also (177)).

For the sake of concreteness, we show below three hypothetical correlograms exemplifying the type cases considered.

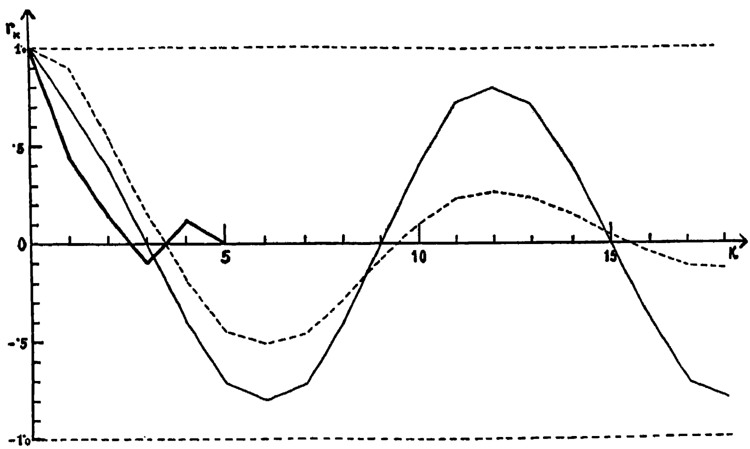

Fig. 6. Correlograms illustrating the schemes of hidden periodicities (thin line), linear autoregression (broken line), and moving averages (thick line).

The correlograms in this figure are based on the following parameters. In the case of hidden periodicities, the correlogram has been derived from formula (46), where we have chosen $s = C_1 = 1$; $D^2(\eta) = \cdot 125$; $\lambda_1 = \pi/6$. The case of linear autoregression is illustrated by a correlogram of type (243), having taken $C = \sqrt{\cdot 8}$, $\lambda_1 = \pi/6$. The moving average correlogram, finally, has been obtained from (250), putting $h = 4$; $b_1 = \cdot 7$, $b_2 = \cdot 4$, $b_3 = -\cdot 3$, $b_4 = \cdot 2$

Because of the different behaviour of the autocorrelation coefficients in the schemes mentioned, it may be expected that we would obtain useful suggestions by inspecting the empirical correlogram when searching for an adequate scheme to be applied to an observational time series. For this reason, the construction of an empirical correlogram is taken as the starting point in the following applications.[16]

It should be observed that the correlogram construction by formula (13) involves a relatively small amount of numerical computa-

tion. No trigonometrical or other mathematical tables are required. Another definite advantage is that the correlogram is obtained directly from the statistical data, without any preceding preparation of the material. Accordingly, the empirical correlogram seems particularly well suited for serving as a first indicator of which type of scheme to apply to the data.

If the empirical correlogram suggests a scheme of hidden periodicities, the next step would be to construct a periodogram for a more detailed analysis of possible periodicities in the material under investigation.

Next, if the correlogram suggests a scheme of linear autoregression, our first problem is to find a scheme (101) such that the corresponding hypothetical correlogram will fit the empirical one. The chief difficulty is to derive suitable values for the coefficients (a) — when having arrived at a set of coefficients (a), the corresponding autocorrelation coefficients will be uniquely determined by the system (221—2), and the residuals $\bar{\eta}_t$ by the relations (280). It is further a desideratum that these residuals be as small as possible. Having seen above that these problems are more intricate than emphasized in earlier studies of the graph of serial coefficients, it will be found that an empirical autoregression analysis as proposed by G. U. Yule (1927) will be useful in this connexion.

Finally, it may happen that the empirical correlogram will suggest a scheme of moving averages. As far as I know, the problem of fitting this scheme to observational data has not been attacked in earlier literature. A fundamental problem in this sphere was formulated by Prof. H. Cramér in his 1933 Course, viz. to find a moving average with a prescribed correlogram. It will be seen that the relation (256) gives a starting point for attacking this problem.

Having now given a preliminary survey, the details of the methods will be discussed when presenting the results of their application. The present section will be concluded by attaching some remarks of general scope concerning the limitation of the methods outlined.

In time series analysis, significance problems are extremely intricate. In dealing with serial coefficients, we have i. a. to pay regard to the fact that their magnitude is conditioned by the size of the statistical masses to which they refer. For a discussion of this point reference is made to Appendix B in the 1st edition of this book (see also p. 110).

The following applications aim only at illustrating the qualitative

differences between various hypothetical models. Consequently, all questions about the significance and the interpretation of the quantitative results fall outside the scope of this study, and again an explicit warning is given against attaching importance to the numerical values found for the parameters of the different models fitted to the observational data.

However, even when limiting the program to an analysis of the qualitative structure of a phenomenon as described by a time series, we must be cautious when interpreting the results. Stating the case briefly: The results will to some extent be conditioned by the methods used in the analysis. Of course, in point of principle the situation is the same in all applications of hypothetical models to empirical data. Let us dwell a moment on some circumstances which are peculiar to time series analysis, especially as based on the correlogram and the autoregression methods.

The correlogram sums up the autocorrelation properties of a time series, and the autoregression analysis, too, is based solely on the autocorrelation coefficients. We conclude that neither method is able to distinguish between different schemes with coincident autocorrelation coefficients. Having already seen examples of this when dealing with the scheme of moving averages (cf. section 26), further examples are readily obtained by using non-linear operations in the construction of stationary processes. For instance, letting $\{\eta(t)\}$ represent a purely random process, it is evident that

$$\xi(t) = \eta(t) \cdot \eta(t-1) \tag{286}$$

will define a stationary process $\{\xi(t)\}$. Assuming that $E[\eta(t)] = 0$, a short calculation will show that the process $\{\xi(t)\}$ is non-autocorrelated.

Thus, if we have found a hypothetical scheme that fits well to an empirical correlogram, it is perfectly possible that there are other schemes which yield an equally close approximation. When it is necessary to choose between different schemes, it may happen that theoretical arguments will speak in favour of one of the schemes. As exemplified in the applications, the schemes of linear regression often seem plausible from theoretical viewpoints, at least as a first approximation. On the other hand, a rational choice between different schemes may be alternatively based on an examination of other structural properties of the time series than its

serial coefficients. Such lines of research, however, fall outside of the program of the present study.

Another point which must be kept in mind has special reference to the autoregression analysis. Having in sections 19 and 20 subjected a general stationary process to an autoregression analysis, the investigation resulted in a canonical form for the process considered, viz. a decomposition in two mutually non-correlated processes, each of a structure readily comprehended in respect to certain fundamental properties. Now, even if a complete parallel to this analysis could be carried through when dealing with empirical data — which of course is impossible, one reason being the necessity of dealing with only a finite number of observations — it is not certain that the autoregression analysis would be an adequate method of research. In fact, in point of principle an autoregression analysis can reveal only linear interrelations between the elements in a time series. For instance, the implicit relation

$$\xi(t) = \eta(t) + p \cdot \xi^2(t-1), \tag{287}$$

where $\{\eta(t)\}$ is purely random, $P[|\eta(t)| < \frac{1}{4}] = 1$, and $|p| < 1$, defines a stationary process $\{\xi(t)\}$; a *linear* autoregression analysis would here give rise to an infinite sequence of residuals, and to a canonical representation which is more complicated than the simple relation (287).

The above argument shows that there is a certain risk of overestimating the outcome of a linear autoregression analysis. When proceeding to residuals of higher order, more parameters are introduced, and it may be that a simpler, possibly non-linear approach would give better results. As mentioned before, the use of non-linear methods does not fall within the scope of this study.

31. Some applications of the scheme of moving averages.

In economic theory, a great deal of interest has recently been paid to the schemes of linear regression. However, the discussion of the advantages of the new ideas over the hypothesis of strict periodicity seems to have been carried on exclusively by general theoretical argumentation, without attempting to fit the recommended schemes directly to observational data. Considering, in particular,

the scheme of moving averages, this has not so far as I know been tried on empirical time series in other fields of scientific research either. When selecting statistical material in order to give applications of the previous analysis, I chose economic time series for one reason because of the lack pointed out. In connexion with the account of these applications given in the sequel, we shall touch upon some related lines of economic research where the previous developments seem to yield proper tools for a deeper analysis.

As mentioned in section 9, J. BARTELS (1935) has found »quasi-persistent periodicity» in certain geophysical time series. A periodogram analysis here being inadequate, these series invite an application of the schemes of linear regression. Thanks to their at once flexible and simple construction, these schemes often seem plausible also *a priori*. A few arguments on this line will be touched upon in the sequel when discussing certain geophysical and other phenomena which from a theoretical viewpoint might be interpreted by means of the schemes of linear regression.

The series of yearly wheat prices in Western Europe 1518—1869 compiled by SIR W. BEVERIDGE (1921) was chosen for my earliest experiment in applying the correlogram method. The purpose being to apply a stationary scheme, the analysis was concerned with BEVERIDGE's trend-free index of fluctuations (p. 449 ff.). In order to avoid changes in the structure of the index, the analysis was restricted to the last hundred data. An account of the analysis follows.

Having made the inconsequential modification of reducing the BEVERIDGE index by 100 units, the time series investigated is given in col. (2) of table 7. The first 15 serial coefficients obtained from this material with the use of formula (13) read as follows.

Table 6. Serial coefficients of the BEVERIDGE *wheat price index 1770—1869.*

$\bar{r}_1 = \cdot 614$,	$\bar{r}_2 = \cdot 090$,	$\bar{r}_3 = -\cdot 156$,	$\bar{r}_4 = -\cdot 115$,	$\bar{r}_5 = -\cdot 006$,
$\bar{r}_6 = \cdot 003$,	$\bar{r}_7 = -\cdot 006$,	$\bar{r}_8 = -\cdot 116$,	$\bar{r}_9 = -\cdot 166$,	$\bar{r}_{10} = -\cdot 102$,
$\bar{r}_{11} = \cdot 033$,	$\bar{r}_{12} = \cdot 084$,	$\bar{r}_{13} = -\cdot 011$,	$\bar{r}_{14} = \cdot 021$,	$\bar{r}_{15} = \cdot 136$.

The correlogram based on these coefficients is shown in fig. 7.

It is seen that $\bar{r}_1$ is rather large, and that all of the following serial coefficients are lying in the interval $-\cdot 17 < \bar{r}_k < \cdot 14$, i. e. rather close to zero. To my eye, the correlogram definitely suggests

a scheme of moving averages. Accordingly, the next step in the analysis will be to search for a moving average with autocorrelation coefficients approximating the serial coefficients under investigation. As mentioned before, this problem is due to H. CRAMÉR (see p. 123).

Quite generally, the problem before us may be stated as follows. A set of numbers $u_1, u_2, \ldots, u_h$ being given, does there exist a moving average (249) with autocorrelation coefficients r_k such that $r_k = u_k$ for $1 \leq k \leq h$? If the answer is in the affirmative, we know from section 26 that there in general will exist a finite group of moving

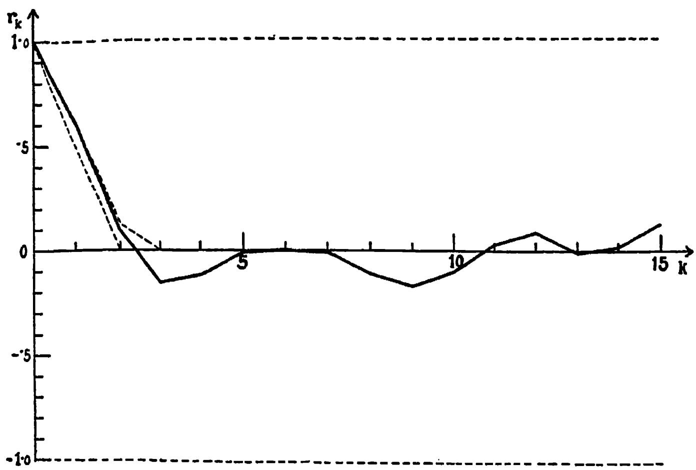

Fig. 7. Correlogram of the BEVERIDGE *wheat price index 1770—1869 (thick line), and hypothetical correlograms corresponding to the approaches* (292) *and* (294) *(broken lines).*

averages with the prescribed autocorrelation coefficients, and we are also in possession of a direct method for determining the coefficients (b) of these moving averages.

Paying regard to the relation (256), we conclude that if there exists a moving average (249) satisfying our conditions, we must have

$$(288) \quad u(x) = u_h x^h + u_{h-1} x^{h-1} + \cdots + u_1 x + 1 + \frac{u_1}{x} + \cdots + \frac{u_{h-1}}{x^{h-1}} + \frac{u_h}{x^h} =$$

$$= \frac{1}{K^2}(x^h + b_1 x^{h-1} + \cdots + b_{h-1} x + b_h)\left(b_h + \frac{b_{h-1}}{x} + \cdots + \frac{b_1}{x^{h-1}} + \frac{1}{x_h}\right).$$

If x_0 is a root of the equation $u(x) = 0$, then $1/x_0$ is another

root. It follows that the substitution $z = x + x^{-1}$ will transform $u(x)$ into a real polynomial in z of order h, say $v(z)$. Let us write

$$v(z) = v_0 z^h + v_1 z^{h-1} + \cdots + v_{h-1} z + v_h. \tag{289}$$

The successive calculation of $v_0, v_1, \ldots$ from the coefficients u_i needs no comment.

It is evident that if z is a root of $v(z) = 0$, two roots of $u(x) = 0$ will be obtained from the equation

$$P(x, z) = x^2 - zx + 1 = 0. \tag{290}$$

The roots of this equation are given by

$$\frac{z}{2} \pm \sqrt{\frac{z^2}{4} - 1}. \tag{291}$$

The product of the roots being unity, we conclude that unless both the roots are of modulus unity, one of them is situated inside, and the other outside the unit circle.

Denoting conjugate complexity by an asterisk, we know that if z is a complex root of $v(z) = 0$, another root reads z^*. Further, if $P(x, z) = 0$ has the roots x and $1/x$, it is evident that $P(x, z^*) = 0$ has the roots x^* and $1/x^*$. In that case one of the real polynomials $(x - x_1)\ (x - x_1^*)$ and $[x - (x_1)^{-1}] \cdot [x - (x_1^*)^{-1}]$ must be a factor in the polynomial

$$b(x) = x^h + b_1 x^{h-1} + \cdots + b_{h-1} x + b_h$$

appearing in (288).

In case $v(z) = 0$ presents a real root, say z_0, we must distinguish two cases. If $|z_0| \geq 2$, both x_1 and x_2 are seen to be real. Keeping in mind that either $x - x_1$ or $x - x_2$ is a factor in $b(x)$, we conclude that this case corresponds to the real roots of the equation $b(x) = 0$.

On the other hand, if $|z_0| < 2$, we know from (291) that x_1 and x_2 are conjugate complex, and of modulus unity. The factors $(x - x_1)$ and $(x - x_2)$ being complex, we conclude that both of them must be contained in $b(x)$. Since one zero of $u(x)$ corresponds to one zero of $v(z)$, this is impossible unless z_0 is a root of even multiplicity of $v(z) = 0$.

After these remarks, the following theorem demands no explanation.

Theorem 12. A necessary and sufficient condition that there exists a moving average (249) with autocorrelation coefficients r_k equalling u_k for $1 \leq k \leq h$ is that the auxiliary polynomial $v(z)$ defined by (289) has no zero z_0 of odd multiplicity in the real interval $-2 < z_0 < 2$.

If this condition is satisfied, the sequences (b) sought for will be given by the coefficients in the real polynomials $b(x)$ satisfying the relation (288). In full agreement with the analysis in section 26, we conclude further from the above discussion that there are at most 2^h such sequences (b), and that the polynomials $b(x)$ may be written on the form $(x - x_1)(x - x_2) \ldots (x - x_h)$, where — denoting by $z_1, z_2, \ldots, z_h$ the roots of $v(z) = 0$ — the real or complex quantity x_1 is a root of $P(x, z_1) = 0$, and x_2 is a root of $P(x, z_2) = 0$, etc.

Returning to the Beveridge index of wheat prices, we shall give a few applications of the method outlined above.

In the correlogram $\bar{r}_k$ (see fig. 7), the small deviations from zero for $k > 1$ might perhaps be looked upon as pure chance products. Thus we are led to investigate whether there exists a moving average $\eta(t) + b_1 \eta(t-1)$ with autocorrelation coefficient r_1 equalling ·614. Putting $h = 1$, and $u_1 = {}$·614, and following the general method, we get $u(x) = {}$·$614\,x + 1 + {}$·$614\,x^{-1}$, and $v(z) = {}$·$614\,z + 1$. Since the root $-1/$·$614 = -1$·63 of $v(z) = 0$ is lying in the critical interval $-2 < z < 2$, we conclude from theorem 12 that there exists no moving average with $r_1 = {}$·614 and $r_k = 0$ for $k > 1$.

A short reflection shows that in all moving averages of type $\eta(t) + b_1 \eta(t-1)$ we have $-$·$5 \leq r_1 \leq$ ·5. It is further evident that there is only one average of this type such that $r_1 = {}$·5, viz.

$$(292) \qquad \zeta(t) - m = \eta(t) + \eta(t-1),$$

Consequently, this average — the correlogram of which is shown in fig. 7 — will yield the closest fit to the prescribed value $\bar{r}_1 = {}$·614. If we rest satisfied with the simple average (292), we have to interpret the deviations between the serial coefficients $\bar{r}_k$ given on p. 151 and the values $r_1 = {}$·5, $r_2 = r_3 = \cdots = 0$ as due to chance. If a better fit is desired, averages involving more parameters (b) must be used. A few examples of this will be given.

As is readily verified, the approach $u_1 = {}$·614, $u_2 = {}$·090 gives $u(x) = {}$·$090\,x^2 + {}$·$614\,x + 1 + {}$·$614\,x^{-1} + {}$·$090\,x^{-2}$, and $v(z) = {}$·$090\,z^2 + {}$ $+ {}$·$614\,z + {}$·820. The roots of $v(z) = 0$ being $z_1 = -1$·82 and

$z_2 = -5{\cdot}00$, it follows from theorem 12 that there does not exist any moving average with the autocorrelation coefficients prescribed. However, a very slight modification in u_2 will suffice to remove z_1 from the critical interval. In fact, expressing that a root (291) shall equal -2, we get $(\frac{1}{2}\bar{r}_1 \cdot u_2^{-1} - 2)^2 = (\frac{1}{2}\bar{r}_1 \cdot u_2^{-1})^2 - (1 - 2u_2)u_2^{-1}$, which gives $u_2 = \bar{r}_1 - \frac{1}{2} = {\cdot}114$.

The function $v(z)$ corresponding to $u_1 = {\cdot}614$, $u_2 = {\cdot}114$ reads ${\cdot}114 z^2 + {\cdot}614 z + {\cdot}772 = 0$, and we get $z_1 = -2{\cdot}000$, $z_2 = -3{\cdot}386$. In order to prepare the construction of the corresponding sequences (b), we solve $P(x, -2) = x^2 + 2x + 1 = 0$, which gives the double root $x = -1$, and $P(x, -3{\cdot}386) = x^2 + 3{\cdot}486\, x + 1 = 0$, which gives the real roots $x = -{\cdot}3269$, and $x = -3{\cdot}0591$. It follows that there exist two binomials $b(x)$ which satisfy the conditions laid down, viz. $b_1(x) = (x+1)(x + {\cdot}3269) = x^2 + 1{\cdot}3269\, x + {\cdot}3269$, and $b_2(x) = (x+1)(x + 3{\cdot}0591) = x^2 + 4{\cdot}0591\, x + 3{\cdot}0591$.

Using the terminology introduced in section 26, the binomial $b_1(x)$ gives rise to a regular moving average, viz.

(293) $$\zeta_1(t) - m = \eta(t) + 1{\cdot}3269\, \eta(t-1) + {\cdot}3269\, \eta(t-2).$$

Since there is only one more process in the group (ζ) of averages with the same correlogram as $\zeta_1(t)$, it is evident that this one is symmetrical with $\zeta_1(t)$, and thus given by

$$\zeta_2(t) - m = {\cdot}3269\, \eta(t) + 1{\cdot}3269\, \eta(t-1) + \eta(t-2).$$

In full agreement with the general theory, $\zeta_2(t)$ can alternatively be derived from $b_2(x)$ by multiplying the coefficients by K_1/K_2, taking for K_1^2 the sum of the squared coefficients in $b_1(x)$, and similarly for K_2^2 — a short calculation will verify that $K_1/K_2 = {\cdot}3269$, etc.

To check the coefficients (b) obtained, it is sufficient to compute the autocorrelation coefficients of $\zeta_1(t)$ from the general formula (250). As it should be, we find $r_1 = {\cdot}614$, $r_2 = {\cdot}114$.

Observing that the increase in u_2 from ${\cdot}090$ to ${\cdot}114$ has brought on a decrease in z_1 from $-1{\cdot}82$ to $-2{\cdot}00$, and an increase in z_2 from $-5{\cdot}00$ to $-3{\cdot}06$, it is seen that the parameters are very susceptible to variations in the initial u-values. Another example of this is given by the fact that a second slight increase in u_2 will make the roots z_1 and z_2 coincide. As follows from (291), this

will occur when u_2 satisfies the relation $(\cdot 307)^2 = u_2(1 - 2\,u_2)$, i.e. when $u_2 = \cdot 1260$.

Using the new value $u_2 = -\cdot 1260$, and taking as before $u_1 = \bar{r}_1 = \cdot 614$, we get $v(z) = \cdot 1260\,z^2 + \cdot 614\,z + \cdot 7480$. The roots of $v(z) = 0$ read $z_1 = z_2 = -\cdot 614 / \cdot 252 = -2\cdot 4365$, and those of $P(x_1, -2\cdot 4365) = 0$ are $x = -\cdot 5225$ and $x = -1\cdot 9140$. Consequently, the binomial $b(x)$ corresponding to the regular one among the moving averages sought for is simply $(x + \cdot 5225)^2 = x^2 + 1\cdot 0450\,x + \cdot 2730$. Thus the regular moving average reads

$$(294) \qquad \zeta_1(t) - m = \eta(t) + 1\cdot 0450\,\eta(t-1) + \cdot 2730\,\eta(t-2).$$

Because of the symmetry, it follows that the moving average corresponding to $b(x) = (x + 1\cdot 9140)^2$ is given by

$$(295) \qquad \zeta_2(t) - m = \cdot 2730\,\eta(t) + 1\cdot 0450\,\eta(t-1) + \eta(t-2).$$

The group (ζ) in this case consists of three processes. The remaining one is obtained from $b(x) = (x + \cdot 5225)(x + 1\cdot 9140) = x^2 + 2\cdot 4365\,x + 1$. A short calculation gives for this process

$$(296) \quad \zeta_3(t) - m = \cdot 5225\,\eta(t) + 1\cdot 2730\,\eta(t-1) + \cdot 5225\,\eta(t-2).$$

Checking the calculations, we find that the autocorrelation coefficients of each of the processes (294)—(296) are given by $r_1 = \cdot 6140$, $r_2 = \cdot 1260$. The correlogram of the group (294)—(296) is shown in fig. 7.

The appreciable effect of even small changes in u_2 is evident from the above. Comparing the regular averages (293) and (294), it is seen that the increase in u_2 from $\cdot 114$ to $\cdot 126$ has caused a decrease in b_1 from $1\cdot 3269$ to $1\cdot 0450$, and a decrease in b_2 from $\cdot 3269$ to $\cdot 2730$.

The following example shows in detail how the method works when $v(z) = 0$ has complex roots. Let us start from the values $u_1 = \cdot 60$, $u_2 = \cdot 09$, $u_3 = -\cdot 15$, $u_4 = -\cdot 10$, which are seen to approximate very closely the first four serial coefficients in the index under investigation. With but little reduction we obtain $-10^3\,v(z) = 10\,z^4 + 15\,z^3 - 49\,z^2 - 105\,z - 62$. Solving $v(z) = 0$, we get $z_1 = -2\cdot 1272$, $z_2 = 2\cdot 5103$, $z_3 = -\cdot 9415 + \cdot 5240\,i$, $z_4 = -\cdot 9415 - \cdot 5240\,i$. Concluding from theorem 12 that there exists a group of moving averages with the prescribed correlogram, a short reflection shows that the group consists of 8 processes.

Solving $P(x, z_1) = 0$, we get $x = -\cdot 7013$, and $x = -1\cdot 4259$, while $P(x, z_2)$ gives $x = \cdot 4966$ and $x = 2\cdot 0137$. The roots of $P(x, z_3) = 0$ were found to be $x = -\cdot 3381 - \cdot 6679\, i$, and $x = -\cdot 6034 + + 1\cdot 1919\, i$. According to the discussion preceding theorem 12, the roots of $P(x, z_4) = 0$ are obtained by replacing i by $-i$ in the roots of $P(x, z_3) = 0$.

Writing

$$B(x) = (x + \cdot 3381 - \cdot 6679\, i)(x + \cdot 3381 + \cdot 6679\, i) = x^2 + + \cdot 6762\, x + \cdot 560402,$$

the regular moving average will be obtained from $b(x) = (x + \cdot 7013)(x - \cdot 4966) \cdot B(x) = x^4 + \cdot 8809\, x^3 + \cdot 3505\, x^2 - \cdot 1208\, x - - \cdot 1952$, and reads

$$(297) \quad \eta(t) + \cdot 8809\, \eta(t-1) + \cdot 3505\, \eta(t-2) - \cdot 1208\, \eta(t-3) - - \cdot 1952\, \eta(t-4).$$

Squaring the coefficients, we get the sum $K^2 = 1\cdot 95153$.

According to the general analysis, a second average with the same correlogram will be delivered by $b_1(x) = (x + 1\cdot 4259)(x - \cdot 4966) \cdot \cdot B(x) = x^4 + 1\cdot 6055\, x^3 + \cdot 4807\, x^2 + \cdot 0420\, x - \cdot 3968$. The sum of squared coefficients is $K_1^2 = 3\cdot 967917$, which gives $K/K_1 = \cdot 701304$. Multiplying the coefficients in $b_1(x)$ by this factor, we get the coefficients in the corresponding moving average. This is found to be

$$\cdot 7013\, \eta(t) + 1\cdot 1259\, \eta(t-1) + \cdot 3371\, \eta(t-2) + \cdot 0294\, \eta(t-3) - - \cdot 2783\, \eta(t-4).$$

A third moving average in the group is seen to be obtained from $(x + \cdot 7013)(x - 2\cdot 0137) \cdot B(x)$. Proceeding as before, we find for the corresponding process

$$\cdot 4966\, \eta(t) - \cdot 3159\, \eta(t-1) - \cdot 8637\, \eta(t-2) - \cdot 8395\, \eta(t-3) - - \cdot 3930\, \eta(t-4).$$

In the same way, the polynomial $(x + 1\cdot 4259)(x - 2\cdot 0137) \cdot B(x)$ yields a fourth process in the group, viz.

$$\cdot 3483\, \eta(t) + \cdot 0308\, \eta(t-1) - \cdot 9432\, \eta(t-2) - \cdot 7909\, \eta(t-3) - - \cdot 5604\, \eta(t-4).$$

The four remaining averages in the group considered correspond to the complex roots $x = -\,{\cdot}6034 \pm 1{\cdot}1919\,i$ of $u(x) = 0$. Due to the symmetry, these processes may be obtained directly from the four processes above by reversing the order of the coefficients. For instance, the regular average (297) gives

$$-\,{\cdot}1952\,\eta(t) - {\cdot}1208\,\eta(t-1) + {\cdot}3505\,\eta(t-2) + {\cdot}8809\,\eta(t-3) + \eta(t-4).$$

The above computations have been checked by verifying that the autocorrelation coefficients of the four processes equal the prescribed values $r_1 = {\cdot}6000$, $r_2 = {\cdot}0900$, $r_3 = -\,{\cdot}1500$, $r_4 = -\,{\cdot}1000$.

Having now exemplified the construction of sets (b) belonging to moving averages (249) with correlograms approximating that of the Beveridge wheat price index, we shall postpone the discussion of the results arrived at until we have made a few applications of the inversion formula (200) and the relation (260).

If $\{\zeta(t)\}$ is a regular moving average (249), the primary process $\{\eta(t)\}$ will be given either by (200) or by (255), the latter formula corresponding to the exceptional case when the characteristic equation of the difference relations satisfied by the coefficients (a) presents roots of modulus unity. This characteristic equation is nothing else than the equation $b(x) = 0$ used in the general method exemplified above. This method being based on the calculation of the roots of the equation mentioned, we are in a position to point out directly which of the formulae (200) and (255) to apply in the different examples, and to carry the analysis further on the basis of the general developments in section 26.

Returning first to the approach (292), we have to apply (255), for the root $x = -1$ of $b(x) = 0$ is of modulus unity. Proceeding as in the illustration 1 of section 26, we get

$$(298)\qquad \eta(t) = \lim_{1>a\to 1}\,[\zeta(t) - m - a\,(\zeta(t-1) - m) + a^2\,(\zeta(t-2) - m) - \cdots]$$

By construction, the second approach (293) is also such that a root of $b(x) = 0$ is of modulus unity.

In order to give an application of (200), we proceed to the approach (294). According to theorem 8, the coefficients (a) will satisfy the difference relation $a_k + 1{\cdot}0450\,a_{k-1} + {\cdot}2730\,a_{k-2} = 0$. The characteristic equation being $b(x) = (x + {\cdot}5225)^2 = 0$, it follows from section 6 that a_k may be written on the form $a_k = (A + k \cdot B) \cdot \cdot (-\,{\cdot}5225)^k$. The constants A and B may be obtained from the

initial values $a_0 = 1$, $a_1 = -1{\cdot}0450$ (cf. (97) and (197)). A short calculation gives $A = B = 1$, and hence $a_k = (1 + k)(-{\cdot}5225)^k$. In applying formula (200), we have to insert this expression, and to replace $\{\zeta(t)\}$ by $\{\zeta_1(t) - m\}$, where $m = E[\zeta_1(t)]$. Observing that $\sum_0^\infty a_k = (1{\cdot}5225)^{-2} = {\cdot}4314$, we get the inversion formula

$$(299)\quad \eta(t) = -{\cdot}4314\, m + \zeta_1(t) - 2({\cdot}5225)\,\zeta_1(t-1) + \\ + 3({\cdot}5225)^2\,\zeta_1(t-2) - 4({\cdot}5225)^3\,\zeta_1(t-3) + \cdots$$

It is seen that the approach (297) gives a formula of the same type, but, of course, with more complicated coefficients (a).

Denoting the BEVERIDGE index by $\bar{\zeta}_t$, and assuming that $\bar{\zeta}_t$ is a sample series of the moving average $\{\zeta_1(t)\}$ given by (294), formula (299) may be used for deriving a series $\bar{\eta}_t$ which corresponds to this hypothesis. Identifying m with the average $\bar{m}$ of the index series $\bar{\eta}_t$, we get

$$(300)\quad \bar{\eta}_t = -{\cdot}4314\,\bar{m} + \bar{\zeta}_t - 1{\cdot}0450\,\bar{\zeta}_{t-1} + {\cdot}8190\,\bar{\zeta}_{t-2} - {\cdot}5706\,\bar{\zeta}_{t-3} + \cdots$$

The sum of the 100 index fluctuations $\bar{\zeta}_t$ given in table 7 being -28, we get $\bar{m} = -{\cdot}28$. Inserting this, and using the index deviations given in col. (1), formula (300) has given the series $\bar{\eta}_t$ presented in col. (2). Having put $a_k = 0$ for $k > 12$, the first 12 $\bar{\eta}_t$-values are partly based on index fluctuations not given in the table.

Apart from the constant $\bar{m} = -{\cdot}28$, the moving average (294) as performed on col. (2) must reproduce col. (1). Thus the values $\bar{\eta}_t$ obtained may be checked by the simple identity

$$(301)\qquad \bar{\zeta}_t - \bar{m} = \bar{\eta}_t + 1{\cdot}0450\,\bar{\eta}_{t-1} + {\cdot}2730\,\bar{\eta}_{t-2}.$$

Because of the large number of terms in (300) and (301), there will sometimes be a deviation amounting to ·2 or ·3.

In analogy with the above, the series $\bar{\eta}_t$ corresponding to the regular approach (292) may be obtained by a limit passage based on the relation (255). Having exemplified in illustration 1 of section 26 the limit procedure for deriving a primary series $\bar{\eta}_t$, it is seen that in point of principle no complications will be met. Accordingly, we shall not dwell further on the exceptional cases when $b(x) = 0$ presents a root of modulus unity.

Starting from the hypothesis that a given time series $\bar{\zeta}_t$ is a sample series of a regular moving average (249), we have above

Table 7. BEVERIDGE *wheat price index fluctuations (col. 1), and hypothetical primary series $\bar{\eta}_t$ (col. 2).*

Year	(1)	(2)	Year	(1)	(2)	Year	(1)	(2)	Year	(1)	(2)
1770	31	23·9	1795	30	17·6	1820	−16	·8	1845	15	19·6
71	36	9·3	96	− 5	−27·2	21	−24	−19·3	46	39	19·9
72	19	3·0	97	−16	7·8	22	−23	− 2·8	47	−10	−35·9
73	6	·5	98	−13	−13·5	23	−29	−20·5	48	−20	12·3
74	5	3·9	99	20	32·2	24	−29	− 6·4	49	−26	−28·8
75	−12	−15·9	1800	39	9·3	25	−31	−18·3	1850	−22	5·0
76	−16	− ·2	01	17	− 1·2	26	−18	3·2	51	−14	−11·1
77	− 6	− 1·2	02	5	4·0	27	− 7	− 5·1	52	5	15·5
78	−13	−11·4	03	− 6	− 9·6	28	14	18·7	53	38	25·1
79	−21	− 8·5	04	25	34·2	29	3	−14.5	54	41	10·8
1780	−13	− ·7	05	14	−18·9	1830	10	20·7	55	38	20·1
81	−12	− 8·6	06	− 2	8·6	31	5	−12·3	56	7	−16·7
82	− 6	3·5	07	− 7	−10·6	32	−18	−10·6	57	−18	− 5·8
83	− 6	− 6·9	08	− 6	3·1	33	−20	− 5·4	58	−19	− 8·1
84	− 8	− 1·3	09	− 6	− 6·0	34	−22	−13·3	59	− 3	7.4
85	−15	−11·4	1810	4	9·8	35	−18	− 2·5	1860	16	10·7
86	−16	− 3·4	11	40	31·8	36	−12	− 5·6	61	7	− 6·0
87	− 7	·0	12	21	−14·5	37	2	8·7	62	− 8	− 4·4
88	8	9·2	13	−4	2·9	38	17	9·6	63	−21	−14·5
89	8	− 1·4	14	− 4	− 2·7	39	7	− 5·2	64	−19	− 2·4
1790	−14	−14·8	15	30	32·3	1840	− 5	− 1·9	65	− 6	·7
91	−22	− 5·9	16	78	45·2	41	1	4·8	66	19	19.3
92	−13	− 2·6	17	26	−29·7	42	− 8	−12·1	67	18	− 1·9
93	−15	−10·5	18	− 6	13·1	43	−12	− ·3	68	− 7	− 9·8
94	3	14·9	19	−14	−19·3	44	− 8	− 4·1	69	2	13·1

exemplified a general method for deriving the corresponding primary series $\bar{\eta}_t$. We are now in a position to solve the analogous problem under the hypothesis that $\bar{\zeta}_t$ is a sample series of a non-regular average (cf. p. 129). After the detailed treatment of the regular case, it will be sufficient to deal with the non-regular averages very briefly.

Let the characteristic equation of a non-regular moving average $\{\zeta(t)\}$ be given by (cf. (258))

$$b(x) = \frac{K}{K^{(i)}}(x^h + b_1^{(i)} x^{h-1} + \cdots + b_{h-1}^{(i)} x + b_h^{(i)}) =$$

$$= \frac{K}{K^{(i)}}(x^p + A_1 x^{p-1} + \cdots + A_{p-1} x + A_p)(x^q + B_1 x^{q-1} + \cdots +$$
$$+ B_{q-1} x + B_q) = 0,$$

where $p + q = h$, and where the roots of $x^p + A_1 x^{p-1} + \cdots + A_p = 0$ are lying inside, and those of $x^q + B_1 x^{q-1} + \cdots + B_q = 0$ outside the unit circle.

The coefficients A_i and B_i being real, two real sequences p_i and q_i will be defined by the systems (cf. (97))

$$\begin{cases} p_n + A_1 p_{n-1} + \cdots + A_{n-1} p_1 + A_n = 0, & n = 1, 2, \ldots, p-1; \\ p_n + A_1 p_{n-1} + \cdots + A_{p-1} p_{n-p+1} + A_p p_{n-p} = 0, & n \geq p; \end{cases}$$

$$\begin{cases} B_q q_n + B_{q-1} q_{n-1} + \cdots + B_{q-n+1} q_1 + B_{q-n} = 0, & n = 1, 2, \ldots, q-1; \\ B_q q_n + B_{q-1} q_{n-1} + \cdots + B_1 q_{n-q+1} + q_{n-q} = 0, & n \geq q; \end{cases}$$

$p_0 = q_0 = 1$. It is seen that the sum

$$\alpha(t) = \zeta(t) + p_1 \zeta(t-1) + p_2 \zeta(t-2) + \cdots$$

will be convergent. Paying regard to certain evident relations between the coefficients A_k, B_k, and $b_k^{(i)}$, it follows further

$$\alpha(t) = K \cdot [\eta(t) + B_1 \eta(t-1) + \cdots + B_q \eta(t-q)] / K^{(i)}. \tag{302}$$

Thus prepared, let us form the sum

$$\beta(t) = \alpha(t) + q_1 \alpha(t+1) + q_2 \alpha(t+2) + \cdots.$$

Observing that the roots of the equation

$$B_q x^q + B_{q-1} x^{q-1} + \cdots + B_1 x + 1 = 0$$

are of modulus less than unity, it follows without difficulty that this sum also is convergent, and that $\beta(t) = B_q \cdot K \cdot \eta(t-q) / K^{(i)}$. Since $B_q \neq 0$, we conclude that, apart from a constant factor, the repeated linear operation performed will yield $\{\eta(t)\}$ in terms of the non-regular moving average considered.

Comprehending the double transformation, we get

$$\{\eta(t)\} = \frac{K^{(i)}}{B_q \cdot K} \cdot \sum_{n=-\infty}^{\infty} c_n \cdot \{\zeta(t + q + n)\}, \tag{303}$$

where $c_0 = 1 + p_1 q_1 + p_2 q_2 + \cdots$, and

$$\begin{cases} c_n = p_n + p_{n+1} q_1 + p_{n+2} q_2 + \cdots, & n > 0, \\ c_{-n} = q_n + q_{n+1} p_1 + q_{n+2} p_2 + \cdots, & n > 0. \end{cases}$$

The generalization to the case when $b(x) = 0$ presents roots on the periphery of the unit circle is, of course, straightforward.

In order to give an explicit application of the relation (303), let us consider the group (ζ) belonging to the regular average (294) studied in detail before. As to (295), we get by analogy from (299), and in full agreement with (303),

$$(304) \qquad \eta(t-2) + {\cdot}4314\, m = \zeta_2(t) - 2\,({\cdot}5225)^2 \zeta_2(t+1) + \\ + 3\,({\cdot}5225)^3 \zeta_2(t+2) - \cdots$$

Considering the remaining average (296) in the group, the coefficients appearing in (303) are seen to reduce to $p_n = q_n = (-{\cdot}5225)^n$. Hence $c_n = p^{|n|}/(1-p^2)$, $p = -{\cdot}5225$, $n \gtreqless 0$. Since $K/K^{(t)}$ is nothing else than the factor of $\eta(t)$ in (296), and $B_q = -1/{\cdot}5225$, we get

$$(305) \qquad \eta(t-1) = \sum_{n=-\infty}^{\infty} [\zeta_3(t+n) - m] \cdot p^{|n|}/(1-p^2), \quad p = -{\cdot}5225.$$

By means of the formulae (304) and (305), it is possible to derive explicitly the two series $\bar{\eta}_t$ corresponding to the hypotheses that the Beveridge index is a moving average (295) or (296) respectively. The calculations running as in the regular case, no detailed illustration will be given. Of course, the last $\bar{\eta}_t$ values can be calculated only approximately. It should be observed that a complete check may be based on identities similar to (301).

If a series $\bar{\eta}_t$ corresponding to a certain process in a group (ζ) has been derived, it will sometimes be possible to arrive at the primary series, say $\bar{\eta}_t^*$, corresponding to another average in the group by means of a simpler procedure than that based on (303). For instance, representing by $\bar{\eta}_t$ and $\bar{\eta}_t^*$ the primary series corresponding to (294) and (296) respectively, the relation (260) gives

$$-\bar{\eta}_t = p\,\bar{\eta}_t^* - (1-p^2)\,\bar{\eta}_{t-1}^* - p(1-p^2)\,\bar{\eta}_{t-2}^* - p^2(1-p^2)\,\bar{\eta}_{t-3}^* - \cdots$$

where $p = -{\cdot}5225$, and hence $p\,\bar{\eta}_t^* - \bar{\eta}_{t-1}^* = p\,\bar{\eta}_{t-1} - \bar{\eta}_t$. The series $p\,\bar{\eta}_{t-1} - \bar{\eta}_t$, say $\bar{\zeta}_t^*$, being readily derived from the series $\bar{\eta}_t$ in table 7, we get the simple relation (cf. (305))

$$\bar{\eta}^*_{t-1} = -\bar{\zeta}^*_t + p\,\bar{\zeta}^*_{t+1} - p^2\,\bar{\zeta}^*_{t+2} + p^3\,\bar{\zeta}^*_{t+3} - \cdots$$

Having now given an account of some applications of the general theory of moving averages to the BEVERIDGE wheat price index, we shall next attach some remarks of general scope to certain points in the analysis. Let us in the first place touch upon the problem of testing the results derived under a hypothesis of moving averages.

In a process $\{\zeta(t)\}$ of moving averages as given by (249), a characteristic feature is that $\zeta(t)$ is independent of $\zeta(t-h-k)$ for $k>0$.

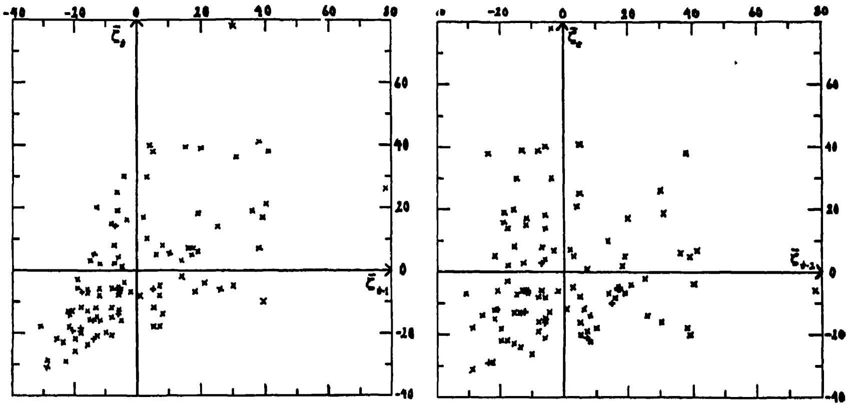

Fig. 8. BEVERIDGE *wheat price index,* $\bar{\zeta}_t$. *The scatters* $(\bar{\zeta}_t, \bar{\zeta}_{t-1})$ *(left), and* $(\bar{\zeta}_t, \bar{\zeta}_{t-2})$ *(right).*

Thus, (249) will form no adequate basis for the analysis of a time series $\bar{\zeta}_t$ unless the scatters $(\bar{\zeta}_t, \bar{\zeta}_{t-h-1})$, $(\bar{\zeta}_t, \bar{\zeta}_{t-h-2})$, etc. approximate distributions of independent variables. On this line it would, of course, be possible to develop different types of tests. — Figure 8 contains the scatters $(\bar{\zeta}_t, \bar{\zeta}_{t-1})$ and $(\bar{\zeta}_t, \bar{\zeta}_{t-2})$ belonging to the BEVERIDGE data given in table 7. Figure 8 (left) clearly shows that $\bar{\zeta}_t$ and $\bar{\zeta}_{t-1}$ must be considered interdependent. On the other hand, the scatter $(\bar{\zeta}_t, \bar{\zeta}_{t-2})$ seems to permit us to look upon $\bar{\zeta}_t$ and $\bar{\zeta}_{t-2}$ as independent, a circumstance speaking in favour of an average of the simple type $\bar{\zeta}_t = \bar{\eta}_t + b\,\bar{\eta}_{t-1}$.

In starting from a specified scheme of moving averages, we are in a position to derive a hypothetical value for any characteristic of the series $\bar{\zeta}_t$ and of the corresponding series $\bar{\eta}_t$ calculated as indicated above. In point of principle, every such characteristic as

compared with the corresponding empirical one will give a basis for testing the hypothetical set-up.

Of course, when a scheme with correlogram approximating the empirical one is chosen, there will automatically be an agreement in the main between certain hypothetical and empirical characteristics. For instance, the serial coefficients of $\bar{\eta}_t$ will approximate zero, since the deviations from this hypothetical value will be due only to the differences between the hypothetical and the empirical correlogram. The situation is the same with regard to the ratio between the variances of $\bar{\eta}_t$ and $\bar{\zeta}_t$. Considering e. g. the approach (294), the hypothetical value of this quotient is $1/(1 + 1{\cdot}0450^2 + {\cdot}2730^2) = {\cdot}462$. On the other hand, the variances of the series $\bar{\eta}_t$ and $\bar{\zeta}_t$ being $213{\cdot}7$ and $383{\cdot}6$ respectively, the empirical quotient equals ${\cdot}557$.

By construction, the moving averages in a group (ζ) will present the same correlogram and the same variance. We conclude from the above that in point of principle the autocorrelation properties of the corresponding series $\bar{\eta}_t$ will give us no criterion for distinguishing which of the processes should be preferred. Expressing the argument in other words, the deviations in $\bar{r}_k(\bar{\eta}_t)$ and $\bar{D}(\bar{\eta}_t)$ from the hypothetical values will depend solely on the differences between the hypothetical and empirical correlograms of $\bar{\zeta}_t$, and these differences are exactly equal for all processes in the group (ζ).

The forecasts based on the hypothesis of moving averages disclose an interesting aspect of the test problem. Denoting by $F_t[\zeta(t+k)]$ the forecast over k time units based on the development up to the time point t, the general formula (213) gives in the approach (294) the following forecasts

$$(306) \qquad F_t[\zeta(t+1)] = \bar{m} + 1{\cdot}0450\,\bar{\eta}_t + {\cdot}2730\,\bar{\eta}_{t-1};$$
$$F_t[\zeta(t+2)] = \bar{m} + {\cdot}2730\,\bar{\eta}_t;$$

and $F_t[\zeta(t+k)] = \bar{m}$ for $k > 2$. In particular, taking $t = 1811$, table 7 shows that $\bar{\eta}_t = 31{\cdot}8$, $\bar{\eta}_{t-1} = 9{\cdot}8$. Keeping in mind that $\bar{m} = -{\cdot}28$, we get $F_t[\zeta(t+1)] = 35{\cdot}6$, $F_t[\zeta(t+2)] = 2{\cdot}4$, while the actual path of the index runs through $\bar{\zeta}_{t+1} = 21$, $\zeta_{t+2} = -4$.

Considering on the other hand the approach (295), which belongs to the same group as (294), the forecast formulae corresponding to (306) would read $F_t[\zeta(t+1)] = \bar{m} + 1.0450\,\bar{\eta}_t + \bar{\eta}_{t-1}$; $F_t[\zeta(t+2)] = = \bar{m} + \bar{\eta}_t$. However, in this case we cannot derive $\bar{\eta}_t$ and $\bar{\eta}_{t-1}$

from the observed values $\bar{\zeta}_t, \bar{\zeta}_{t-1}, \bar{\zeta}_{t-2}, \ldots$ Now, the autoregression analysis as developed in Chapter II yields a linear forecast which is valid for any stationary process with finite dispersion, and for which the squared deviation from the future path is of minimum expectation. Writing $\{\eta^{(i)}(t)\}$ for the residual process of the stationary process considered, and paying regard to the results arrived at in section 26, this forecast will in the present case reduce to

$$(307) \qquad F_t[\zeta(t+k)] = b_k\, \eta_t^{(i)} + b_{k+1}\, \eta_{t-1}^{(i)} + \cdots + b_h\, \eta_{t-h+k}^{(i)},$$

where the sequence (b) is identical with the coefficients in the corresponding regular average. According to the general theory, $\{\eta^{(i)}(t)\}$ is obtained simply by subjecting $\{\zeta(t)\}$ to the same linear operation which gives the primary process $\{\eta(t)\}$ in terms of the regular average.

In taking the squared deviation as the measure of the efficiency of a forecast, we conclude from the above that the different hypothetical averages in a group (ζ) will give rise to exactly the same forecast series $F_t[\zeta(t+1)]$, $F_t[\zeta(t+2)]$, etc. A simple illustration of this is given by the group (294)—(296). Taking e.g. the process (295) for a hypothetical basis, we have in the first place to form the series $\bar{\eta}_t^{(i)}$. According to the general analysis, this is identical to col. (2) in table 7. Applying next the general relation (307), and keeping in mind that the coefficients (b) coincide with those in the regular average (294), it follows that the resulting forecasts will equal those previously obtained on the basis of the regular average (294).

Expressing the situation in other words, we have found that the different averages in a group (ζ) are equivalent in view of those aspects of the general test problem hitherto considered. Since the indeterminateness is due to our having dealt with only the autocorrelation properties of the hypothetical models, we have to use other types of methods for distinguishing between the different averages in a group. Generally speaking, these tests should examine to what degree the elements $\bar{\eta}_t$ and $\bar{\eta}_{t+k}$ resulting from a special hypothesis might be considered not only uncorrelated but also independent. The different averages in a group giving rise to different primary series $\bar{\eta}_t$, we must therefore compare in detail the multi-dimensional scatters $(\bar{\eta}_t, \bar{\eta}_{t-1}, \ldots, \bar{\eta}_{t-h})$. For the sake of concreteness, the scatter $(\bar{\eta}_t, \bar{\eta}_{t-1})$ obtained from col. (2) in table 7 is

shown in fig. 9. To my eye, the scatter forms no very good approximation to a distribution of independent variables. In performing different tests of this nature, it may of course occur that no series $\bar{\eta}_t$ will give satisfactory results — perhaps one of the scatters examined will suggest instead a non-linear scheme, e.g. a moving average performed on the non-autocorrelated process $\{\xi(t)\}$ defined by (286). However, it would be beyond the frame of this study to enter upon further details concerning the non-linear schemes and the non-linear methods required for distinguishing between the different moving averages belonging to the same group.

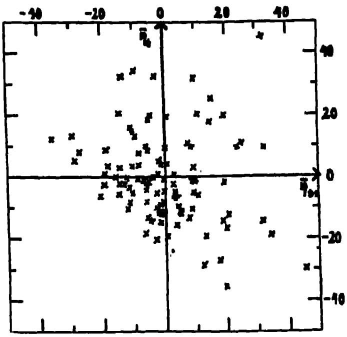

Fig. 9. BEVERIDGE *wheat price index. The scatter* $(\bar{\eta}_t, \bar{\eta}_{t-1})$ *belonging to a hypothetical primary series* $\bar{\eta}_t$.

As repeatedly emphasized in earlier sections, the applications accounted for in the present study do not aim at reaching definitive, quantitative results. Accordingly, no attempts will be made to test the significance of the parameters arrived at under the different hypotheses dealt with. Hence it is out of the question to draw quantitative comparisons with earlier investigations of the wheat price data analysed. As is well known, SIR W. BEVERIDGE (1922) has subjected his index to an extensive periodogram analysis, while G. U. YULE (1926) has illustrated certain new correlogram methods by the use of the BEVERIDGE index numbers. However, it will be illuminating to take up some points of these investigations for a qualitative comparison with the moving average approach.

In his presentation of the wheat price data, SIR W. BEVERIDGE (1921) gives two series of index numbers, the trend present in the first series being removed in the second. In contradistinction to BEVERIDGE and to the present writer, G. U. YULE (1926) works on the first series. In his search for hidden periodicities, YULE modifies the correlogram method because of the trend present in the original data. Since YULE's approach takes into consideration the differences of the series analysed, it might be interpreted as an application of certain non-stationary or evolutive processes of the homogeneous type $\{\xi(t)\}$ defined by (192). The investigation thus following a quite different line of research, no further comment is called for in the present connexion.

The periodogram method being based on the assumption of hidden periodicities, its use will result in a hypothetical scheme additively built up by a functional element $y(t)$ consisting of a number of superposed harmonics, and a random element or »error» $\eta(t)$. Since a single harmonic involves three parameters, the total number of parameters in a scheme of hidden periodicities amounts to thrice the number of harmonics superposed. Specifying the values for these parameters, we can compute the functional element $y(t)$. Deducting $y(t)$ from the original data, say $\bar{\zeta}_t$, we obtain the corresponding path of the random component, $\bar{\eta}_t = \bar{\zeta}_t - y(t)$.

A common standard measure of the efficiency of the periodogram analysis is obtained by dividing the variance of the errors $\bar{\eta}_t$ by the variance of the original data $\bar{\zeta}_t$. This quotient, say $\bar{\varkappa}^2$, is always below unity, and the closer to zero the quotient, the more the functional element $y(t)$ will »explain» of the series under analysis. For instance, the harmonic corresponding to the largest ordinate in the periodogram 1545—1844 given by SIR W. BEVERIDGE ((1922), p. 438) will leave an »error» $\bar{\eta}_t$ with variance amounting to 91 % of the variance of the wheat price index. The harmonic in question is of period $p \sim 15{\cdot}225$.

On the other hand, a scheme of moving averages (249) is built up by means of a random variable $\eta(t)$, here called primary, and a set of coefficients or parameters (b). After having chosen numerical values for the parameters (b), the general analysis in the present section provides a method for deriving from the original series $\bar{\zeta}_t$ the corresponding path $\bar{\eta}_t$ of the primary variable.

Even in the case of moving averages, the quotient $\bar{\varkappa}^2$ between the variances of $\bar{\eta}_t$ and $\bar{\zeta}_t$ may be taken as a standard measure of the efficiency of the analysis. The hypothetical value for $\bar{\varkappa}^2$ being in this case $\varkappa^2 = 1/(1 + b_1^2 + \cdots + b_h^2)$, it is seen that the larger the coefficients (b), the less is $\varkappa^2$, and the better the result of the analysis. For instance, considering the $\bar{\eta}_t$-values in table 7 derived from the approach (294), we have, as already mentioned, $\bar{\varkappa}^2 = {\cdot}557$, $\varkappa^2 = {\cdot}462$. The small values obtained seem rather satisfactory, but it must be remembered that the analysis has been restricted to the last 100 BEVERIDGE data 1770—1869. Thus, although the approach (294) involves only two parameters, it might perhaps be unfair to the scheme of hidden periodicities to compare directly with the $\bar{\varkappa}^2$-value ${\cdot}91$ derived from the whole series under the hypothesis of one harmonic component.

An examination of the forecasts delivered will show clearly the thorough-going difference between the scheme of hidden periodicities and the scheme of moving averages.

In a scheme of hidden periodicities of type (39), the forecast curve is identical to the functional element $y(t)$, i.e. the sum of harmonics superposed (cf. p. 59). The hypothetical model thus will provide a definite functional forecast over the infinite future. The expected squared deviation between the actual development and the forecast is independent of the period forecasted over, and amounts to $D^2(\eta)$.

A scheme of moving averages gives a quite different type of forecast. In fact, considering a moving average (249) ranging over $h+1$ time units, it is only the forecasts over the next h observations that are effective — the forecasts beyond h time units are trivial, and reduce to the average of the original data. According to the general analysis, the different averages in a group (ζ) will give rise to the same forecasts, viz.

$$F_t[\zeta(t+k)] = \bar{m} + b_k\,\bar{\eta}_t + b_{k+1}\,\bar{\eta}_{t-1} + \cdots + b_h\,\bar{\eta}_{t-h+k},$$

where the coefficients (b) are those appearing in the regular average used as the basis for deriving the primary series $\bar{\eta}_t$. The hypothetical value for the squared deviation from the actual development being (cf. (218))

$$(1 + b_1^2 + b_2^2 + \cdots + b_{k-1}^2)\cdot D^2(\eta),$$

it is seen that the efficiency of the prognosis will decrease gradually as the period forecasted over is extended.

Especially in view of economic time series, the type of forecast delivered by the scheme of moving averages seems *a priori* more realistic, seems to correspond better to what might be reasonably possible to find out from the past development. Further, considering the forecasts over a short period, the prognosis given by the scheme of moving averages is, as a rule, rather efficient. In my opinion, this is a circumstance of central importance, for often the main interest is concentrated upon the prognosis concerning the near future.

As to the harmonic components suggested by a periodogram analysis, these cannot always be interpreted in the light of what is otherwise known of the phenomena under investigation. Lacking

other evidence, the periodicities thus will stand out as quite isolated results of the analysis. As pointed out by SIR W. BEVERIDGE ((1922), p. 438), this is the case with the period of length 15·225 years suggested in his periodogram mentioned above. — Against this background, the scheme of moving averages seems more fertile. Let us dwell a moment on this point.

In modern economic-statistical research work, a prominent line of approach is the regression analysis of time series (cf. e.g. C. F. Roos (1934)). Denoting a set of time series by $\bar{\zeta}_t$, $\bar{\eta}'_t$, $\bar{\eta}''_t, \ldots, \bar{\eta}_t^{(n)}$, a simple type of approach reads

$$(308) \qquad \bar{\zeta}_t = c_1 \bar{\eta}'_t + c_2 \bar{\eta}''_t + \cdots + c_n \bar{\eta}_t^{(n)} + \bar{\varepsilon}_t,$$

where $\bar{\varepsilon}_t$ stands for the residual in the representation of $\bar{\zeta}_t$ as linearly correlated with the n variables $\bar{\eta}_t^{(i)}$. A generalized approach is obtained by replacing $\bar{\eta}_t^{(i)}$ by $\bar{\eta}_{t-k_i}^{(i)}$, where the constant k_i represents the *lag* of the series $\bar{\eta}_t^{(i)}$ behind the series $\bar{\zeta}_t$. The well-known concept of *distributed lag* implies a further generalization, which in the simplest case $n=1$ leads to an approach of type

$$(309) \qquad \bar{\zeta}_t = b_0 \bar{\eta}_t + b_1 \bar{\eta}_{t-1} + \cdots + b_h \bar{\eta}_{t-h} + \bar{\varepsilon}_t.$$

For instance, as an initial approximation we might represent a wheat price (i.e. $\bar{\zeta}$) in the year t as linearly correlated with wheat crops (i.e. $\bar{\eta}$) in the years $t, t-1, \ldots, t-h$. According to the theory of supply and demand, we might expect that in this case the dominant regression coefficients b_i would be negative.

Disregarding the residual $\bar{\varepsilon}_t$, and interpreting in the terminology of the present study, the series $\bar{\zeta}_t$ as given by (309) is seen to be nothing else than a moving average performed on the primary series $\bar{\eta}_t$. Consequently, a hypothetical model corresponding to the approach (309) will be obtained by adding two independent processes, viz. a moving average $\{b_0 \eta(t) + b_1 \eta(t-1) + \cdots + b_h \eta(t-h)\}$ and a residual process $\{\varepsilon(t)\}$.

Having seen that the concept of moving average may be attached directly to the concept of *distributed lag*, the theory of moving averages as developed in the present study seems to disclose new aspects of lag problems, and to suggest methods for a deeper analysis in this field of research. A few remarks on this line will follow.

The idea underlying the general methods applied in the present section is that the autocorrelation properties of a time series $\bar{\zeta}_t$ will reveal whether the process of moving averages is an adequate type of hypothetical model for $\bar{\zeta}_t$. The circumstances of importance in this connexion are (A) that in such an analysis no other time series are taken into consideration, and (B) that a specified hypothesis of moving averages (it should be observed that the different averages in a group (ζ) must be examined separately!) will determine a hypothetical primary series $\bar{\eta}_t$. Accordingly, independently of the first stage of the analysis we can compare such a hypothetical series $\bar{\eta}_t$ with other time series, say $\bar{\eta}_t'$, $\bar{\eta}_t''$, etc., thought of as possibly affecting the series $\bar{\zeta}_t$ examined. For instance, in applying the approach (294) to the Beveridge wheat price index, we have derived the hypothetical series $\bar{\eta}_t$ given in col. (2) of table 7. This series has been calculated merely in order to illustrate in detail how the general inversion formula (200) works when applied to an observational time series, but in case the parameters (b) were significant, the series $\bar{\eta}_t$ might be compared with e.g. some appropriate wheat crop series $\bar{\eta}_t'$. Following the suggestion made in connexion with the approach (309), we might in such a case change simultaneously the signs in the coefficients (b) and the series $\bar{\eta}_t$.

In view of the lines of research suggested above, our theoretical analysis of moving averages calls for generalizations in various directions. Having assumed the primary process $\{\eta(t)\}$ in (249) to be purely random, it would in the first place be of interest to generalize the concept of moving average by removing the restrictions imposed on the primary process $\{\eta(t)\}$. Now, assuming only that $\{\eta(t)\}$ is stationary and of finite dispersion, the autocorrelation coefficients of $\{\zeta(t)\}$ will exist, and be obtained by a straightforward generalization of the relations (250). Considering two time series $\bar{\zeta}_t$ and $\bar{\eta}_t'$ with correlograms $\bar{r}_k(\bar{\zeta})$ and $\bar{r}_k(\bar{\eta}_t')$ respectively, the generalized relations evidently may be used for finding out approximately how a moving average with specified coefficients (b) as performed on $\bar{\eta}_t'$ would transform $\bar{r}_k(\bar{\eta}_t')$. If the transformed correlogram approximates $\bar{r}_k(\bar{\zeta})$, we are led to investigate in detail whether $\bar{\zeta}_t$ approximates a moving average with these coefficients (b), and with $\bar{\eta}_t'$ for primary series. Concluding from theorem 9 that the representations (200) and (303) hold even for the generalized average, this investigation can be performed as suggested in the case of a purely random $\{\eta(t)\}$, viz. by deriving directly from $\bar{\zeta}_t$ the primary series

$\bar{\eta}_t$ corresponding to the coefficients (b), and then correlating or comparing in another way the two series $\bar{\eta}_t$ and $\bar{\eta}'_t$.

Finally, another generalization is introduced when several series $\bar{\eta}'_t$, $\bar{\eta}''_t, \ldots$ are employed for building up the series $\bar{\zeta}_t$ (cf. (308)). Since such an approach falls under the theory of multi-dimensional stochastical processes, a discussion of this case would be out of place in the present study.

In examining the general methods applied to the BEVERIDGE wheat price index, we have laid stress upon the different type of forecast delivered by the scheme of moving averages as compared with the scheme of hidden periodicities. The approach of moving averages seems particularly useful in forecasting over a short period of time, and attaches directly to current forecast methods, in particular the approach of distributed lag. Having in the previous analysis referred throughout to economic-statistical applications, the present section will be concluded by a few remarks concerning the applicability of the scheme of moving averages in other fields of scientific research.

In periodogram analysis of geophysical data — e.g. records of rainfall, water-levels, temperature, terrestrial magnetism, etc. — it is often difficult to interpret the periods suggested as physical realities. The situation being the same as in economic applications, the claiming of such periodicities has been subjected to severe criticism. A fact especially stressed — see e.g. an excellent critical survey by D. BRUNT (1937) — is that these periodicities can explain but a small, often quite insignificant fraction of the variance in the observational data. In view of this, the lines of research based on multi-dimensional regression analysis seem more promising (see e.g. C. W. B. NORMAND (1932)). Aiming at short time forecasts, the realistic hypothesis underlying these investigations is that the phenomenon considered is causally connected with other phenomena by relations involving distributed lags. The theoretical set-up required having been touched upon in our discussion of economic applications, it is seen that the methods suggested by the theory of moving averages might be used also in these fields of research.

For instance, representing the water-level in a lake by a moving average of the rainfall in surrounding districts, and following the method outlined, we are led to compare the water-level correlogram with the rainfall correlogram as transformed by the hypothetical moving average. By the courtesy of my friend B. BRUNO, who

has taken an interest in my studies in time series analysis, I am in a position to give an illustration of this on the basis of correlograms appearing in a forthcoming paper (B. BRUNO (1938)).

Starting from quarterly observations 1807—1936 of the level of *Lake Väner*, 130 yearly data $\bar{\zeta}_t$ were obtained by simple averaging. Using formula (13), BRUNO has derived a set of serial coefficients $\bar{r}_k(\bar{\zeta})$ for each of the periods 1807—1936 and 1871—1930. The resulting correlograms are shown in fig. 10.

If a strict periodicity were present in the material, the two correlograms $\bar{r}_k(\bar{\zeta})$ should rise simultaneously to a maximum with

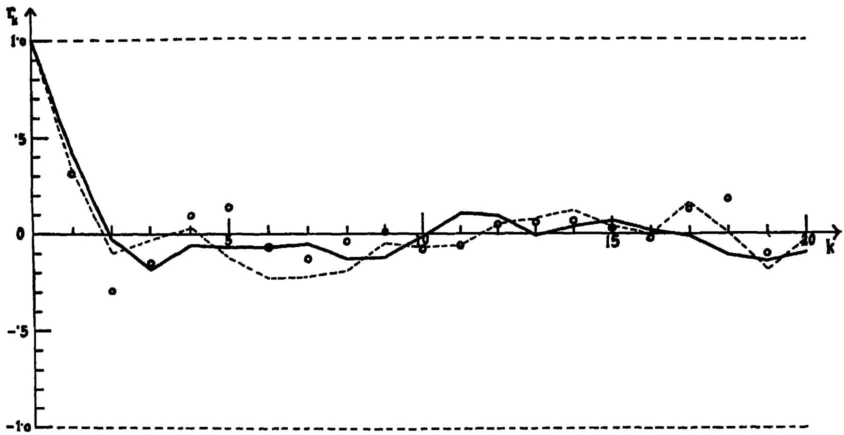

Fig. 10. Correlogram of the level of Lake Väner 1807—1936 (thick line), and 1871—1930) (broken line). Transformed rainfall correlogram (small rings).

abscissa equalling the length of the period. However, the correlograms fluctuating rather independently, no period is suggested in this way. Anyhow, in view of the smallness of the serial coefficients $\bar{r}_k(\bar{\zeta})$ obtained for $k > 1$, a hypothetical period cannot be expected to explain a significant part of the variance in the series $\bar{\zeta}_t$ under analysis.

To my eye, the correlograms suggest more definitely a scheme of moving averages of the simple type $\{\zeta(t)\} = b_0\{\eta(t)\} + b_1\{\eta(t-1)\}$, where $\{\eta(t)\}$ is purely random. In fact, only the coefficient $\bar{r}_1(\bar{\zeta})$ is rather large in both the correlograms, while — as should be the case if the remaining serial coefficients were chance products — those based on the longer period of observation are, on the whole, lying closer to zero. Taking $r_1(\zeta) = \cdot 4$ for the hypothetical value of $\bar{r}_1(\bar{\zeta})$, we have seen in illustration 2 in section 26 (see. p. 131) that the

corresponding group (ζ) of averages is constituted by the two processes $\{\eta(t)\} + \cdot 5\,\{\eta(t-1)\}$ and $\cdot 5\,\{\eta(t)\} + \{\eta(t-1)\}$.

B. Bruno has further constructed the correlogram of a series $\bar{\eta}'_t$ obtained by averaging the yearly rainfall 1867—1936 in four cities in or near the drainage-basin of Lake Väner, viz. Falun, Karlstad, Vänersborg, and Oslo. The correlogram obtained being shown in fig. 11, it is seen that the deviations from zero of $\bar{r}_k(\bar{\eta}')$ are rather irregular and small.

Assuming as a first approach the serial coefficients $\bar{r}_k(\bar{\eta}')$ to be insignificant, we obtain a closed hypothetical model of the two

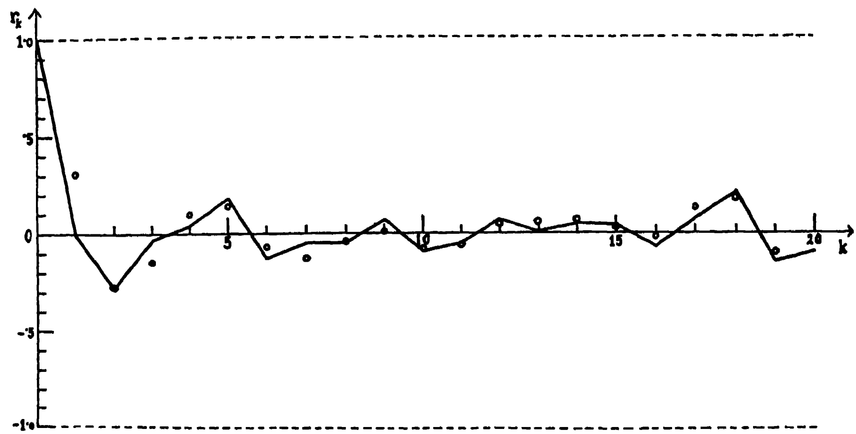

Fig. 11. Correlogram of the rainfall 1867—1936 in the drainage basin of Lake Väner (thick line), and the same correlogram transformed by formula (310) (small rings).

series $\bar{\zeta}_t$ and $\bar{\eta}'_t$ by regarding the rainfall series $\bar{\eta}'_t$ as belonging to a purely random process $\{\eta'(t)\}$, and the water-level series $\bar{\zeta}_t$ as belonging to one of the averages in the group constituted by $\{\eta'(t)\} + \cdot 5\,\{\eta'(t-1)\}$ and $\cdot 5\,\{\eta'(t)\} + \{\eta'(t-1)\}$.

As suggested on p. 170, this simple model might be generalized by cancelling the assumption that $\{\eta'(t)\}$ be purely random. Letting in such a case the autocorrelation coefficients of $\{\eta'(t)\}$ be represented by $r_k(\eta')$, a short calculation shows that those of $\{\zeta'(t)\} = b_0\,\{\eta'(t)\} + b_1\,\{\eta'(t-1)\}$ will be given by

$$(310) \qquad r_k(\zeta') = \frac{(b_0^2 + b_1^2)\cdot r_k(\eta') + b_0\, b_1\,[r_{k+1}(\eta') + r_{k-1}(\eta')]}{b_0^2 + b_1^2 + 2\, b_0\, b_1 \cdot r_1(\eta')},$$

(cf. a related formula given by G. U. Yule (1926), Appendix II). Replacing $r_k(\eta')$ by $\bar{r}_k(\bar{\eta}')$, this formula gives with good approximation

the correlogram of the series $b_0 \bar{\eta}'_t + b_1 \bar{\eta}'_{t-1}$. Putting $b_0 = 1$, and $b_1 = \cdot 6$, we have in this manner obtained the transformed correlogram indicated by small rings in fig. 11.* For direct comparison with the water-level correlograms, the coefficients $\bar{r}_k(\bar{\eta}')$ as transformed by (310) have also been plotted in fig. 10. The parallelism with the correlogram based on the period 1871—1930 is rather encouraging — in 10 cases out of 13 the rings and the coefficients $\bar{r}_k(\bar{\zeta})$ vary in the same direction. Observing that the transformation (310) is symmetrical in respect of the coefficients (b), we are thus led to examine in detail to what extent the water-level $\bar{\zeta}_t$ may be approximated by a moving average of the rainfall $\bar{\eta}'_t$ of the simple form $\bar{\eta}'_t + \cdot 6\, \bar{\eta}'_{t-1}$ or $\cdot 6\, \bar{\eta}'_t + \bar{\eta}'_{t-1}$.

32. Some applications of the scheme of linear autoregression.

Having in Chapter III investigated the process of linear autoregression on the basis of a theory of stochastical difference equations, we shall in the present section give a few applications of this scheme. In doing this, we shall proceed as in the previous section. Choosing an economic time series as our experimental object, we shall first illustrate in detail a general method for determining the parameters when applying a scheme of linear autoregression. The modest purpose being to show how the method works, the tests touched upon in the following discussion will not be applied for examining the significance of the parameters arrived at. In discussing the results, the analysis will instead be focussed on a qualitative comparison with other hypothetical schemes, and with certain related lines of economic-statistical research work. Following up the parallelism with the applications of the scheme of moving averages, this section will be concluded by referring to a few other fields of scientific research which invite an application of the scheme of linear autoregression.

The time series dealt with in the experiments accounted for in the following is the Swedish cost of living index 1830—1913

---

* Similarly, (310) gives the serial coefficients of the series $(\bar{\beta}_t)$ extracted in table 2 in terms of $\bar{r}_k(\bar{a}_t^{(2)})$ as given on p. 50.

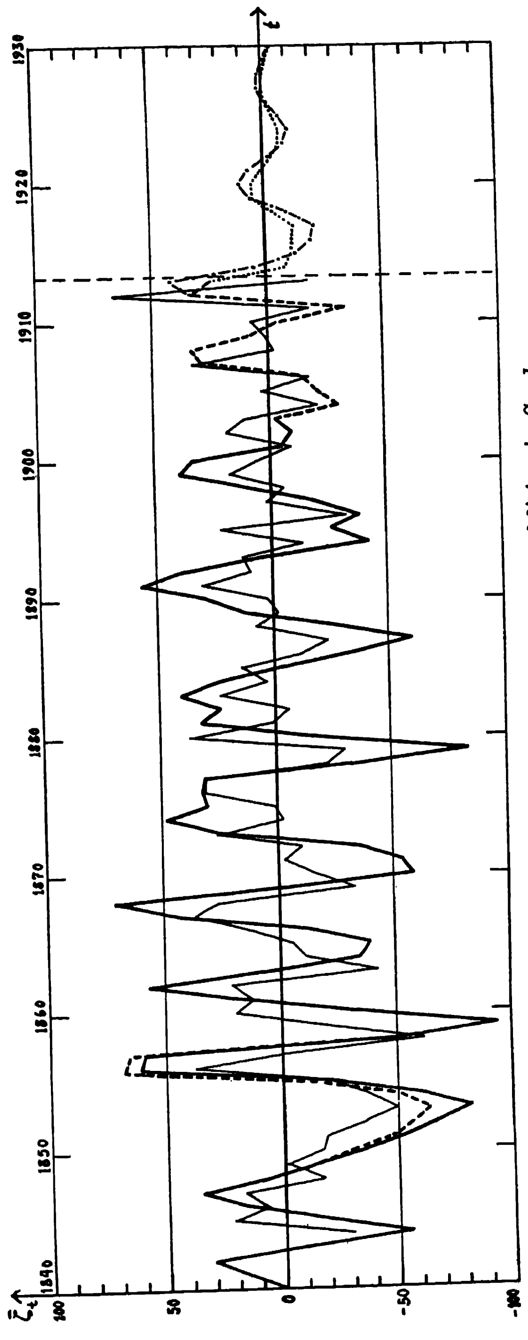

Fig. 12. G. Myrdal *index of the cost of living in Sweden.*
Deviations from Spencer *trend* (——)
Deviations from adjusted trend (— — —)
Hypothetical primary series (——)
Forecast curves (. . . . . . , —.—.—.)

compiled by G. MYRDAL ((1933), Table A, budget b). Since the index presents a marked trend, this had to be removed before starting the analysis (cf. p. 146). Having used the 21-term formula of J. SPENCER for this purpose (see e. g. E. T. WHITTAKER and G. ROBINSON (1926), p. 290 f), the deviations from the graduated index are shown in fig. 12 (thick line). The graduation being disturbed by a rapid rise of the index in the years 1851—57, the values obtained for this period were subjected to a slight adjustment (broken line). Another graduation by hand was used for the years 1904—13 not covered by the SPENCER formula (broken line). In order to avoid decimals, the resulting 84 deviations 1840—1913 were multiplied by 10. These index fluctuations, say $\bar{\zeta}_t$, constitute our experimental series, and are shown in fig. 12. The numerical values of $\bar{\zeta}_t$ are given in col. (1) of table 8.

Table 8. Fluctuations in the G. MYRDAL *cost of living index (col. 1), and hypothetical primary series $\bar{\eta}_t$ (col. 2).*

Year	(1)	(2)	Year	(1)	(2)	Year	(1)	(2)	Year	(1)	(2)
1840	0	·	1860	−33	19·7	1880	−14	36·9	1900	34	5·5
41	16	·	61	21	12.6	81	32	2·1	01	− 5	− 9·1
42	31	·	62	58	20·4	82	24	− 5·9	02	− 9	18·8
43	− 6	·	63	8	−42·4	83	40	23·3	03	− 2	11·2
44	−56	−29·6	64	−35	−11·4	84	24	3.7	04	−30	−21·2
45	−21	21·8	65	−39	− 5·2	85	− 3	14·5	05	−21	3·6
46	16	6·9	66	−13	12·2	86	−35	−11·9	06	−14	−17·6
47	35	16·9	67	43	36·5	87	−60	−23·6	07	29	31·9
48	9	−18·0	68	71	26·6	88	−26	7·1	08	33	− 2·2
49	−13	− 1·6	69	− 1	−33·8	89	12	− 1·9	09	8	·6
1850	−36	−17·0	1870	−59	−16·7	1890	29	2·7	1910	− 5	6·1
51	−50	−19·0	71	−54	− 3·0	91	56	30·4	11	−34	−17·8
52	−58	−36·4	72	−36	−10·5	92	39	8·8	12	31	65·4
53	−64	−49·0	73	25	26·8	93	2	12·2	13	24	−17·4
54	−48	−38·6	74	48	− 3·0	94	−42	−13·3		−198	−47·5
55	−11	−21·1	75	30	1·6	95	−26	21·9			
56	68	38·5	76	32	30·3	96	−38	−32·4			
57	66	− 3·7	77	31	30·2	97	−19	1·9			
58	−48	−62·8	78	−33	−22·6	98	11	− 6·0			
59	−92	−19·3	79	−84	−30·0	99	39	17·9			

Our series $\bar{\zeta}_t$ is seen to reflect clearly the changes between economic expansion and contraction. A certain regularity seems to be present in the movement up and down, but the distance between two adjacent maxima is rather inconstant, varying between some 5 and 10 years. The structure of the fluctuations is summed up in the correlogram obtained from formula (13). This is shown in fig. 14, and the numerical values are given in col. (1) of table 9.

The correlogram looks rather like a simple damped oscillation, say $C \cdot q^k \cdot \cos(\lambda k + \varphi)$. An inspection of the graph shows that in approximating the correlogram by such a function we would have to take the period $p = 2\pi/\lambda$ to be about 7 or 8 years, the phase φ to be approximately vanishing, and $q^7 \sim {}^1/_2$, the latter relation corresponding to a damping of some 50 % in the duration of one period.

According to the theoretical developments in section 25, a process $\{\zeta(t)\}$ of linear autoregression as defined by a relation of type

$$\zeta(t) + a_1 \zeta(t-1) + a_2 \zeta(t-2) = \eta(t) \tag{311}$$

will present a correlogram forming a simple damped harmonic. On the other hand, in a scheme of hidden periodicities, each of the harmonic components will give rise to an undamped harmonic in the correlogram. We conclude that such a correlogram cannot approximate the graph of serial cofficients unless at least two harmonics are superposed. However, such a scheme would involve 6 or more parameters, while only two are required in (311). In seeking for a simple hypothetical scheme with correlogram approximating $\bar{r}_k$ as shown in fig. 14, we are thus led to try first a process of linear autoregression, and firstly one of the simple type defined by (311).

In a general process (219) of linear autoregression, the linear system (222) with coefficients $a_1 \ldots, a_h$ will deliver the autocorrelation coefficients $r_1, \ldots, r_{h-1}$ required for deriving the following coefficients r_h, r_{h+1}, etc. from the difference relations (221). In searching for an adequate scheme (311), we are confronted with the inverse problem, viz. to find a set of coefficients $a_1, \ldots, a_h$ giving rise to a hypothetical process (311) with correlogram approximating a prescribed, empirical correlogram. Now, observing that the relations (221) — (223) are linear also in respect of the coefficients (a), a convenient starting point for attacking the problem before us would be to replace the coefficients r_k in (222) and the

last relation (221) — a system whose determinant, $\Delta(r, n-1)$, is $\neq 0$ — by the prescribed values $\bar{r}_k$, and to solve instead for the coefficients (a). The system yielding our trial set (a) thus will read

$$\text{(312)} \qquad \begin{cases} \bar{r}_1 + \ \ a_1 \ \ + a_2\bar{r}_1 + a_3\bar{r}_2 + \cdots + a_h\bar{r}_{h-1} = 0 \\ \bar{r}_2 + a_1\bar{r}_1 + \ \ a_2 \ \ + a_3\bar{r}_1 + \cdots + a_h\bar{r}_{h-2} = 0 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \bar{r}_h + a_1\bar{r}_{h-1} + a_2\bar{r}_{h-2} + a_3\bar{r}_{h-3} + \cdots + a_h = 0. \end{cases}$$

If the roots of the equation (34) formed by the resulting coefficients (a) are lying in the unit circle, these coefficients will define a process (219) of linear autoregression. By construction, the autocorrelation coefficients $r_1, \ldots, r_h$ of this process will coincide with the prescribed values. The following coefficients r_{h+1}, r_{h+2}, etc. will be obtained from the difference relations (221) formed by the hypothetical coefficients (a).

Having derived the correlogram r_k of the hypothetical process, we are in a position to compare it with the empirical correlogram $\bar{r}_k$. If the fit seems satisfactory, we may carry the analysis further on the basis of the coefficients (a) found, but if the deviations seem too large, an adjustment in the coefficients (a) is called for.

In analogy with the case of moving averages, we obtain according to (219) a primary series

$$\text{(313)} \qquad \bar{\eta}_t = \xi_t - \bar{m} + a_1(\xi_{t-1} - \bar{m}) + \cdots + a_h(\xi_{t-h} - \bar{m})$$

corresponding to our set (a). In other words, the complete hypothetical model will consist of h parameters (a) and the primary series $\bar{\eta}_t$; using these quantities, we can reconstruct the original series ξ_t. In this case, too, we may look upon the quotient $\bar{\varkappa}^2 = \Sigma\,\bar{\eta}_t^2 / \Sigma\,\bar{\xi}_t^2$ as a measure of the efficiency of the analysis — the closer to zero our $\bar{\varkappa}^2$, the less important is the »unexplained» random component, and the greater the efficiency of the approach.

As mentioned in section 29, the method of G. U. Yule (1927) for determining the coefficients (a) in the approach (313) is to minimize $\Sigma\,\bar{\eta}_t^2$. The sum $\Sigma\,\bar{\xi}_t^2$ being constant, this method will make $\bar{\varkappa}^2$ a minimum, and may thus be said to be of maximum efficiency. Having stated this, let it be pointed out explicitly that the Yule method gives a system for determining the coefficients (a), which is equivalent to the system (312) started from above. In fact, minimizing $\Sigma\,\bar{\eta}_t^2$ as defined by (313), we obtain a set of normal equations, where, apart from a constant factor, the coefficients of the

a_i's will obviously approximate the corresponding serial coefficients in the system (312).

The variance of the process $\{\eta(t)\}$ defined by our hypothetical set (a) will be given by formula (220). Of course, this relation holds irrespective of the method used in determining the set (a). On the other hand, choosing the system (312) for determining the coefficients (a), the resulting primary series $\bar{\eta}_t$ will evidently satisfy the parallel relation

(314) $$\bar{D}^2(\bar{\eta}_t) \backsimeq (1 + a_1 r_1 + a_2 r_2 + \cdots + a_h r_h) \cdot \bar{D}^2(\bar{\zeta}_t),$$

where the sign $\backsim$ covers the approximation made in disregarding the first h terms in the series $\bar{\zeta}_t$ for which we cannot calculate corresponding elements $\bar{\eta}_t$. In other words, the variance of the primary series $\bar{\eta}_t$ will approximate the hypothetical value $D^2(\eta)$. It must be kept in mind that this will not always be the case if the trial set (a) is determined otherwise than by (312) (cf. p. 145).

Summing up, the system (312) will give us a set $a_1, \ldots, a_h$ which will minimize the variance of the corresponding residuals $\bar{\eta}_t$. Further, the first h autocorrelation coefficients will coincide with the corresponding serial coefficients. However, the hypothetical correlogram will not always in its whole range yield a good fit to the empirical correlogram. In practice, we can compromise between the two desiderata of obtaining small residuals $\bar{\eta}_t$ and small deviations between the correlograms, and besides try to satisfy the relation $D^2(\eta) \sim \bar{D}^2(\bar{\eta}_t)$. Before discussing this matter, let us see in detail how the method outlined will function when applied to the G. Myrdal cost of living index.

Forming the system (312) for $h=2$, inserting the values $\bar{r}_1 = \cdot 5216$, $\bar{r}_2 = -\cdot 2240$ given in col. (1) of table 8, and solving for the coefficients a_1 and a_2, we obtain $a_1 = -\cdot 8771$, $a_2 = \cdot 6815$. The roots of the characteristic equation $z^2 + a_1 z + a_2 = 0$ being $\cdot 4385 \pm \cdot 6994\, i$, and thus less than unity in modulus, we conclude that the relation

(315) $$\zeta(t) - \cdot 8771\, \zeta(t-1) + \cdot 6815\, \zeta(t-2) = \eta(t)$$

will define a process of linear autoregression. By construction, the first two autocorrelation coefficients of this process $\{\zeta(t)\}$ read $r_1 = \bar{r}_1 = \cdot 5216$, $r_2 = \bar{r}_2 = -\cdot 2240$, while the following coefficients will be obtained recurrently from the difference relation $r_k + \cdot 8871\, r_{k-1} - -\cdot 6815\, r_{k-2} = 0$. The resulting correlogram, which is evidently of the form (243), is shown in fig. 14 (thin line).

Table 9. Serial coefficients $\bar{r}_k$ of the G. MYRDAL *cost of living index (col.* (1))*, *and autocorrelation coefficients r_k belonging to the schemes* (318) (*col.* (2)), *and* (319) (*col.* (3)).

k	(1)	(2)	(3)	k	(1)	(2)	(3)
1	·5216	·5216	·5386	11	—·1533	—·1722	—·2170
2	—·2240	—·2240	—·1460	12	—·2530	—·0218	—·1536
3	—·5811	—·5811	—·5024	13	—·2254	·1065	—·0119
4	—·4626	—·4626	—·5105	14	—·0042	·1311	·1042
5	—·0963	—·0734	—·2320	15	·1883	·0609	.1318
6	·2035	·2749	·1417	16	·1601	—·0383	·0747
7	·3138	·3538	·3458	17	·0723	—·0818	—·0140
8	·2613	·1717	·2902	18	—·0136	—·0630	—·0743
9	·1434	—·0820	·0772	19	—·0067	—·0062	—·0766
10	—·0034	—·2172	—·1321	20	·0342	·0408	—·0327

Comparing with the empirical correlogram, it is seen that the period in the hypothetical correlogram is too short, and that the damping is a little too heavy. According to section 6, the damping factor equals $\sqrt{a_2}$, while the period is given by $p = 2\pi/\lambda$, where $\cos\lambda = -a_1/2\sqrt{a_2}$. Thus, an increase in a_2 will bring on a slighter damping. Further, reducing λ we obtain a longer period. However, as pointed out in the previous discussion, we cannot conclude without further evidence that it will be possible to improve the fit — the coefficients a_1 and a_2 determine also the constant factor and the phase of the damped harmonic, and it might happen that an adjustment in a_1 and a_2 would cause such a change, e. g. in the phase, that the total result of the adjustment would be a poorer fit. Proceeding with the illustration, we shall next examine the total effect of an adjustment.

In the correlogram of the approach (315), the period is found to be 6·22 years, while a good fit would require a period not below 7 years. Reducing the damping by increasing a_2 from ·6815 to ·77, a short calculation will verify that $a_1 = 1·10$ will give a period $p = 7·03$ years. Now, let us examine the approach

$$(316) \qquad \zeta(t) - 1·10\,\zeta(t-1) + ·77\,\zeta(t-2) = \eta(t).$$

* In reading the proofs, a slight error was discovered in the serial coefficients: they should all have been multiplied by a factor amounting to about 1·005.

The correlogram of the process $\{\zeta(t)\}$ thus defined has been calculated from (221) and (222), and is shown in fig. 14 (broken line). Up to r_8 and r_9, the hypothetical correlogram seems to fit rather well. Beyond this point, the fit is less satisfactory, partly because the graph of serial coefficients presents a slow descent to the minimum in $k \sim 12{\cdot}5$, and a rapid rise to the next maximum. This skewness will be recurred to later.

A clear view of the adjustment will be obtained by calculating the roots of the characteristic equations of (315) and (316). In fig. 13, these roots are indicated by small rings. In drawing the conjugate roots nearer to the periphery of the unit circle, the damping has been reduced, while the reduction in the angle λ has elongated the period.

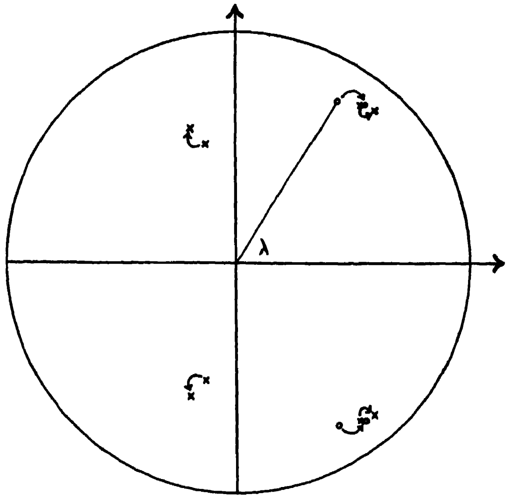

Fig. 13. Adjustment paths in the approaches (315) *(rings), and* (318) *(crosses).*

Each of the approaches (315) and (316) gives rise to a primary series $\bar{\eta}_t$, and we know from a previous remark of general scope that the improvement in the fit of the hypothetical correlogram is obtained at the expense of an increase in the variance of the series $\bar{\eta}_t$. In the approach (315), which is based on a system of type (312), the relation (314) will hold.

Thus, paying regard to (220), and inserting $a_1 = -{\cdot}8771$, $a_2 = {\cdot}6815$, $r_1 = {\cdot}5216$, $r_2 = -{\cdot}2240$, we obtain in this case $\bar{D}^2(\bar{\eta}_t) \sim D^2(\eta) = {\cdot}390\, D^2(\zeta)$. As to the residuals $\bar{\eta}_t$ derived from the approach (316), the development

$$(317) \quad \bar{D}^2(\bar{\eta}_t) = \bar{D}^2(\bar{\zeta}_t + a_1 \bar{\zeta}_{t-1} + a_2 \bar{\zeta}_{t-2}) \sim \sim (1 + a_1^2 + a_2^2 + 2\, a_1 \bar{r}_1 + 2\, a_2 \bar{r}_2 + 2\, a_1 a_2 \bar{r}_1) \cdot \bar{D}^2(\bar{\zeta}_t)$$

will not reduce to (314). Inserting $a_1 = -1{\cdot}10$, $a_2 = {\cdot}77$, $\bar{r}_1 = {\cdot}5216$, $\bar{r}_2 = -{\cdot}2240$, we find in this case $\bar{D}^2(\bar{\eta}_t) \simeq {\cdot}427\, \bar{D}^2(\bar{\zeta}_t)$. In full agreement with the general theory, the adjustment in the coefficients (a)

has reduced the efficiency of the approach. As mentioned in connexion with (314), the hypothetical variance $D^2(\eta)$ will also be affected by the adjustment. Generally speaking, nothing compels $D^2(\eta)$ to follow the variation in $\bar{D}^2(\bar{\eta}_t)$. Actually, in the present case $D^2(\eta)$ varies contrarily to $\bar{D}^2(\bar{\eta}_t)$ — the first two autocorrelation coefficients in the scheme (316) being $r_1 = \cdot 6215$, $r_2 = -\cdot 0864$ (cf. fig. 14), it is readily verified by inserting these values together with $a_1 = -1\cdot 10$, $a_2 = \cdot 77$ in formula (220) that the variance in question will be given by $D^2(\eta) = \cdot 250\ D^2(\zeta)$.

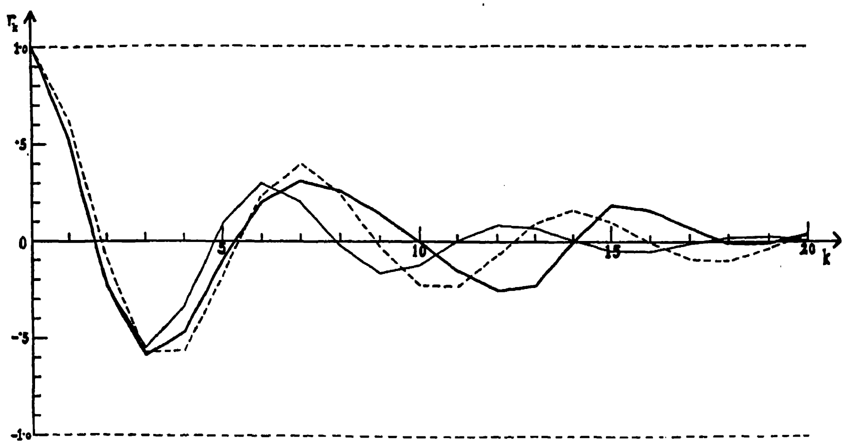

Fig. 14. Correlogram of the G. MYRDAL *cost of living index (thick line), and hypothetical correlogram corresponding to formula (315) (thin line), and formula (316) (broken line).*

Summing up the comparison between the approaches (315) and (316), the hypothetical correlogram fits better in the latter case, but the variance $\bar{D}^2(\bar{\eta}_t)$ is smaller in the former case and coincides with the hypothetical variance $D^2(\eta)$ — in the approach (316) the difference between these variances is rather large. All in all, neither of the schemes seems adequate. In view of other experiments with approaches of the simple type (311), it seems as if we cannot find a satisfactory approach without taking into account more distant elements $\bar{\zeta}_{t-3}$, $\bar{\zeta}_{t-4}$, etc.

Of course, having different desiderata to comply with in applying a scheme of linear autoregression, we could agree on which weights to attach to them, and then take them into consideration simultaneously. In point of principle, it would then be possible to find

out a set (a) forming the best compromise in the sense agreed. However, judging from certain experiments of this kind, what might be gained in this way seems not worth the extensive computations involved. One way or another, the results arrived at in these experiments merit no recital. Accordingly, proceeding to an account of certain experiments with a scheme (219) involving four parameters (a), we shall follow the same line as before.

Taking $h = 4$, and inserting in the system (312) the serial coefficients $\bar{r}_k$ given in table 9, we arrive at the following approach

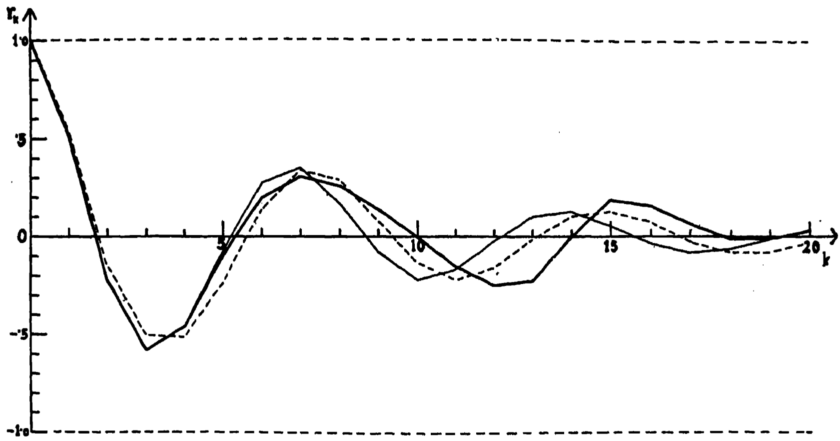

Fig. 15. Correlogram of the G. MYRDAL *cost of living index (thick line), and hypothetical correlogram corresponding to formula (318) (thin line), and formula (319) (broken line).*

$$(318) \qquad \zeta(t) - \cdot 8100\, \zeta(t-1) + \cdot 7452\, \zeta(t-2) - \cdot 0987\, \zeta(t-3) + \\ + \cdot 2101\, \zeta(t-4) = \eta(t).$$

The autocorrelation coefficients of the process $\{\zeta(t)\}$ thus defined have been calculated from (221) and (222). The values found are given in col. (2) of table 9, and plotted in fig. 15 (thin line).

It is seen that the approaches (316) and (318) give rise to almost coincident correlograms, only that here we have by construction $r_k = \bar{r}_k$ for $k \leq 4$. As shown in fig. 13, this conformity is reflected in the characteristic equations — in the present case two of the four roots are lying in the close neighbourhood of the roots belonging to the approach (316). The numerical values found for the roots in the approach (318) read $\cdot 5385 \pm \cdot 6814\, i$, $- \cdot 1335 \pm \cdot 5106\, i$.

In adjusting the approach (318), we are at liberty to move the roots of the characteristic equation in any directions, keeping in mind that complex roots must be conjugate. The dominant component in the correlogram evidently corresponds to the roots $\cdot 5385 \pm \cdot 6814\, i$, say $\varrho \cdot e^{\pm i \lambda}$. The period of this component is nearly 7 years, while the empirical correlogram suggests a somewhat longer period. Now, reducing the angle λ so as to obtain a period equalling 7·5 years, the resulting set of coefficients (a) gave rise to a correlogram with reduced amplitudes. Neutralizing this effect by reducing the damping by means of a slight move towards the periphery of the unit circle, it was found adequate to perform a simultaneous move in the other two roots. As a matter of fact, the deviations between r_k and $\bar{r}_k$ for $k = 1 - 4$ caused by the adjustments mentioned were found to be substantially reduced by moving the root $-\cdot 1335 + \cdot 5106\, i$ in a direction nearly opposite to the adjustment in the root $\cdot 5385 + \cdot 6814\, i$.

Having thus arrived at the roots $\cdot 5888 \pm \cdot 6540\, i$, $-\cdot 20 \pm \cdot 58\, i$, the paths followed are indicated in fig. 13. As is readily verified, the adjusted roots belong to the approach

$$(319) \qquad \zeta(t) - \cdot 7776\, \zeta(t-1) + \cdot 6797\, \zeta(t-2) - \cdot 1342\, \zeta(t-3) + \\ + \cdot 2914\, \zeta(t-4) = \eta(t).$$

The autocorrelation coefficients of this scheme as derived from the system (222) and the relations (221) are given in col. (3) of table 9, and plotted in fig. 15 (broken line). The fit to the empirical correlogram is not very close, but the general shape of the hypothetical correlogram is rather satisfactory. Comparing with the approach (316), which involves only two parameters, it is seen that the improvement bears chiefly upon the period and the coefficient r_1.

Using formula (314), and paying regard to the identities $r_k = \bar{r}_k$, $k \leq 4$, the approach (318) gives $D^2(\eta) = \cdot 371\, D^2(\zeta)$, $\bar{D}^2(\bar{\eta}) = \cdot 371\, \bar{D}^2(\bar{\zeta})$. Comparing with the simple scheme (315), it is seen that the reduction of the factor in $D^2(\eta)$ and $\bar{D}^2(\bar{\eta})$ amounts to only $\cdot 019$. In other words, the introduction of two more parameters has brought on but a slight increase in the efficiency of the approach. However, comparing the adjusted schemes, we find a definite improvement. Proceeding as in (317), we find in the first place that the residuals derived from (319) will satisfy the relation $\bar{D}^2(\bar{\eta}) \simeq \cdot 381\, \bar{D}^2(\bar{\zeta})$ (working directly on the series $\bar{\zeta}_t$ and $\bar{\eta}_t$ given in table 8, we find

$\bar{D}^2(\bar{\eta}_t) = {}^{\cdot}385\,\bar{D}^2(\bar{\zeta}_t)$. The adjustment has thus reduced the efficiency of the approach but very slightly. On the other hand, applying formula (220), we find that the parallel hypothetical relation reads $D^2(\eta) = {}^{\cdot}401\, D^2(\zeta)$. Contrary to the situation in the approach (316), it is seen that $\bar{D}^2(\bar{\eta}_t)$ in the present case approximates $D^2(\eta)$ even after the adjustment.

Perhaps it would be possible to find an adjustment improving the approach (319). However, in view of the above figures, not much can be gained by a continued adjusting. Nor does it seem

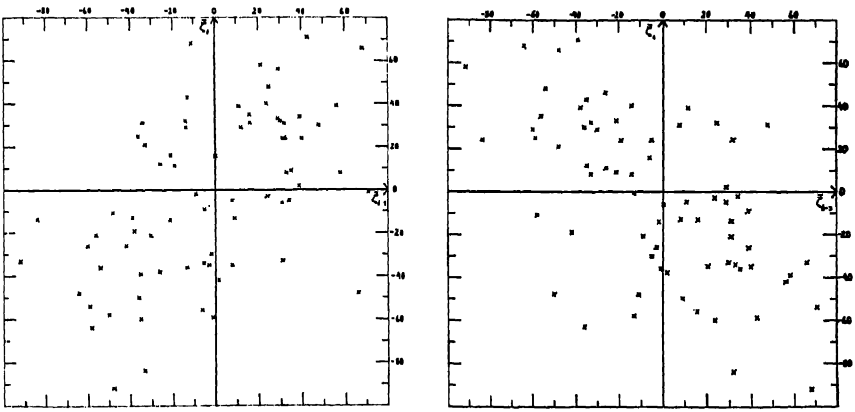

Fig. 16. G. MYRDAL *cost of living index $\bar{\zeta}_t$. The scatters $(\bar{\zeta}_t, \bar{\zeta}_{t-1})$ (left), and $(\bar{\zeta}_t, \bar{\zeta}_{t-3})$ (right).*

as if a real improvement could be secured by enlarging our set (a). Be that as it may, the above examples are sufficient for our purpose of illustrating a general method for deriving a trial set of coefficients (a) in applying a process of linear autoregression, and for performing adjustments in the trial set.

Proceeding to a discussion of the approach (319), the same general viewpoints present themselves as in the case of moving averages. Accordingly, referring to the remarks in the previous section (cf. p. 163), we need draw attention only to a few circumstances which are peculiar to the scheme of linear autoregression.

As indicated by the term proposed, the hypothesis of linear autoregression implies a linear regression upon $\bar{\zeta}_t$ of each of the preceding elements $\bar{\zeta}_{t-1}$, $\bar{\zeta}_{t-2}$, etc. This circumstance having already been employed by G. U. YULE (1927) for testing an approach of this kind (cf. section 29, p. 141), we show in fig. 16 the scatters

$(\bar{\zeta}_t, \bar{\zeta}_{t-1})$ and $(\bar{\zeta}_t, \bar{\zeta}_{t-3})$ of the fluctuations in the MYRDAL index. The deviations from linearity in the connexion between the variables do not seem disturbing.

As in the case of moving averages, different tests of a scheme of linear autoregression may be based on the hypothetical randomness of the primary series $\bar{\eta}_t$. Such tests may, for example, be focussed upon the scatters $(\bar{\eta}_t, \bar{\eta}_{t-k})$. The scatter $(\bar{\eta}_t, \bar{\eta}_{t-1})$ formed by the residuals $\bar{\eta}_t$ given in table 8 is shown in fig 17.

The above mentioned skewness in the correlogram of the G. MYRDAL index (see fig. 14) gives an interesting illustration of the difficulties of testing time series schemes. Is it permitted to look upon the deviation from the hypothetical correlogram as produced by pure chance? An examination of this question must pay due regard to the interdependence of the serial coefficients in a sample series. For instance, a chance deviation in $k = 13$ would perhaps be most often attended by such deviations in the neighbouring coefficients that the total picture would present a skew oscillation. The question of how much weight to attach to deviations of this and similar kinds seems extremely intricate. Perhaps nothing better can be done than to compute a large number of model series correlograms, and then to compare the deviations from the hypothetical curve.

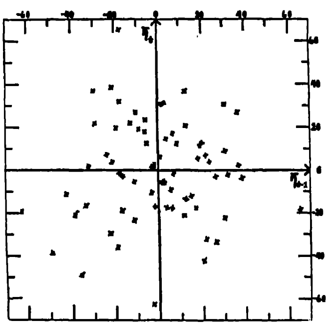

Fig. 17. Scatter $(\bar{\eta}_t, \bar{\eta}_{t-1})$ of a hypothetical primary series of G. MYRDAL'S *index of the cost of living.*

As pointed out in section 24, an approach of linear autoregression will give rise to a forecast curve which forms a damped oscillation of type (33), and satisfies the same linear difference equation as the autocorrelation coefficients. Letting as before $F_t[\zeta(t+k)]$ represent a forecast over k time units, formula (214) shows that $F_t[\zeta(t+k)]$ may be conveniently computed recurrently. For instance, considering the approach (319), we get $F_t[\zeta(t+1)] = (1 + a_1 + a_2 + a_3 + a_4) \cdot \cdot \bar{m} - a_1 \bar{\zeta}_t - a_2 \bar{\zeta}_{t-1} - a_3 \bar{\zeta}_{t-2} - a_4 \bar{\zeta}_{t-3}$, $F_t[\zeta(t+2)] = (1 + a_1 + a_2 + + a_3 + a_4) \cdot \bar{m} - a_1 \cdot F_t[\zeta(t+1)] - a_2 \bar{\zeta}_t - a_3 \bar{\zeta}_{t-1} - a_4 \bar{\zeta}_{t-2}$, etc. In particular, inserting for $t = 1912$ the values of $\bar{\zeta}_t$ given in col (1) of table 8, we find $F_{1912}[\zeta(1913)] = 41{\cdot}4$, $F_{1912}[\zeta(1914)] = 5{\cdot}2$, etc. The forecast curve $F_{1912}[\zeta(1912+k)]$ is shown in fig. 12 (dashes and dots). A check of the first forecast is obtained by deducting

the hypothetical residual (»random shock») $\bar{\eta}_{1913} = -17{\cdot}4$; the result is 24, which equals $\bar{\zeta}_{1913}$. Fig. 12 shows also the forecast curve $F_{1913}\,[\zeta(1913 + k)]$, (dotted line).

The two forecast curves in fig. 12 yield a good illustration of the prognosis situation in an approach of linear autoregression. Firstly, while a forecast $F_t[\zeta(t + k)]$ is often rather efficient for small k-values, the efficiency vanishes asymptotically as k increases (cf. p. 165). Further, as soon as we are in a position to take a new observation $\bar{\zeta}_{t+1}$ into consideration when forming the prognosis, the forecast curve is often substantially modified; how much, will depend on the residual $\bar{\eta}_{t+1} = \bar{\zeta}_{t+1} - F_t[\zeta(t + 1)]$. — Summing up, it is the short forecasts that are efficient. In this respect, we meet the same situation as in the scheme of moving averages, and the same contrast to the scheme of hidden periodicities (cf. p. 168). On the other hand, under special circumstances the oscillations in a scheme of linear regression are nearly functional, viz. nearly strictly periodic — as remarked in discussing the sinusoidal limit theorem of E. Slutsky (cf. p. 120), processes of hidden periodicities can be obtained as limit cases of the schemes of linear autoregression.

As pointed out in section 21, the scheme of linear autoregression constitutes the proper starting point when studying oscillatory mechanisms which are subjected to random impulses. A typical approach of this kind is formed by the *complete systems* as dealt with in several recent economic studies (see e. g. R. Frisch (1933), J. Tinbergen (1937)). A simple example of a complete system is given by

$$(320) \qquad \begin{cases} \xi(t) = c_1\,\zeta(t-1) + \eta'(t), \\ \zeta(t) = d_0\,\xi(t) + d_1\,\xi(t-1) + \eta''(t). \end{cases}$$

For instance, as a first approximation we may take the production volume $\xi(t)$ of a commodity to be linearly correlated with the price $\zeta(t-1)$, and the price $\zeta(t)$ to be linearly built up by the production volumes $\xi(t)$ and $\xi(t-1)$.

In this connexion, the pertinent thing is that a complete system may be reduced to a single relation involving but one of the fundamental variables ξ, ζ, etc. Considering e. g. the simple case (320), we get at once

$$(321) \qquad \zeta(t) + a_1\,\zeta(t-1) + a_2\,\zeta(t-2) = \eta(t),$$

where a_1 and a_2 are constants, and η is linear in the variables η' and η''.

Of course, in order to study a complete system in detail, we must consider stochastical processes in several dimensions, a generalization not in the program of the present study. However, it is evident that theorem 9 applies directly to the reduced relations exemplified by (321). Thus, under general conditions concerning the variables $\eta(t)$ and the constants (a), the variables $\zeta(t)$ form a stationary process. In point of principle, we are in a position to investigate the properties of this process $\{\zeta(t)\}$. Considering e. g. the autocorrelation coefficients, it follows without difficulty that these will satisfy a linear difference equation with a right member.

It is seen that the theory of linear autoregression as developed in sections 24 and 25 covers the case when the complete system reduces to a relation with a purely random right member. We have seen that this analysis has given certain results which cannot be reached by functional methods. For instance, if the left member of the relation (321) is characterized by an intrinsic damped oscillation, and if the damping is too heavy, the tendency to periodicity cannot be distinguished in the mechanism as subjected to random impulses. Another example of this is given by the approach (319), which presents two intrinsic periods. Having already mentioned that one of these equals 7·5 years, a short calculation will verify that the other period is 3·44 years. Now, as shown in fig. 13, the root corresponding to the latter period is lying rather near the centre of the unit circle, which implies that the damping is rather heavy. Thus, even if the period 3·44 is quantitatively reliable and significant — which seems to me rather doubtful — we cannot conclude without further analysis that this period can be found by a periodogram construction or similar methods. Be that as it may, as pointed out before I attach no importance to the quantitative significance of the above analysis of the MYRDAL index.

Starting from explicit assumptions about the relations between the different variables in an economic system, E. LUNDBERG (1937) has examined how the system will develop from hypothetical initial conditions. Since several of the relations assumed are non-linear, his approach may by looked upon as a generalization of the linear systems as exemplified by (320). Now, the analysis of E. LUNDBERG shows that the variables will often present tendencies to diverge

as an economic expansion goes on, tendencies causing tensions which make the economic system instable. Of course, in point of principle it would be possible to apply stochastical methods even in this approach. However, it seems extremely difficult to obtain in this way a non-evolutive scheme for the system considered. Anyhow, in view of the investigations of E. LUNDBERG, the linear approaches earlier discussed seem far from sufficient for giving an adequate hypothetical model for studying the economic cycles in detail. Looking upon the approach (319) from this viewpoint, it might be said that even if this scheme does correctly sum up the main features of the index examined, the interpretation in terms of oscillatory mechanisms which is suggested by such a simple approach cannot possibly be completely realistic.[17]

Having in the previous section mentioned certain geophysical phenomena which invite to studies on the basis of the scheme of moving averages, this section will be terminated by a few suggestions about the wide applicability of the scheme of linear autoregression.

As surveyed in section 29, G. U. YULE (1927) introduces the concept of autoregression in studying the 11 year wave in the sunspot numbers. Further, since the criterion of J. BARTELS (1935) (cf. p. 26) suggests that certain waves in terrestrial magnetism are »quasi-persistent», the construction of this criterion makes us expect that in these cases an autoregression approach will be fruitful. Of course, here the scheme of linear autoregression suggests itself also *a priori*. For instance, let us consider the 27 day wave, which is due to the sunspot intensity and the rotation of sun. The duration of a sunspot often being rather long, the sunspot intensity as observed in a time point t, say ξ_t, must be positively correlated with the intensity 27 days earlier. Having stated this, it seems plausible that a correlogram of terrestrial magnetism will present an oscillation with a period of about 27 days. We must also expect that the oscillation in the empirical correlogram will be damped, and that the degree of damping will depend on the average duration of a sunspot.

It is not difficult to point out other instances where the scheme of linear autoregression seems plausible on theoretical grounds. Even the simple scheme (237) involving but one parameter might often prove useful, at least as a first approximation. For instance, we have already found that the scheme (284) presents a correlogram which fits rather well to the correlogram of air pressure examined

by SIR G. WALKER (1931). In this case, the autoregression obviously may be interpreted as an effect of inertia — generally speaking, we may always expect an autoregression of type (237), and with a positive constant p, when dealing with phenomena characterized by irreversibility. According to formula (239), the expectation corresponding to a long period is in such cases larger than in a purely random series. In other words, there will be a tendency to spurious periodicity, a tendency of the observational time series to present long waves the lengths of which are varying and without physical significance.

Finally, it is evident that the theory of autocorrelation may be applied to the functional transform (21) used by N. WIENER (1930) in the theory of light. Following up this idea, we are led to interpreting the transmission of light as a stationary process. With suitable arrangements about the dispersion of the process, the expression (21) would then correspond to the autocorrelation coefficients, while the function $S(\lambda)$ as given by (23) would reduce to the generating function of the autocorrelation coefficients (cf. section 17). In this approach, the continuous parts of the spectrum of light would correspond to the continuity intervals of the generating function. Perhaps the scheme of linear autoregression might serve as a starting point for investigations on these lines, for according to theorem 11 this scheme presents a continuous generating function, and, as pointed out in section 25 (see p. 120), we may obtain any scheme of hidden periodicities of type (39) as a limit case.

Appendix 1

Notes to the second edition

[1] The names of the various schemes and processes were in the 1st edition introduced by way of tentative suggestions. To some extent they have been accepted in later works [see, for instance, refs. 2, 28, 29, 40], but the terminology is not too well established, and the author thinks it would be to advantage to change some of the terms. For one thing, now that the pioneer works of YULE and SLUTSKY can be viewed at a longer distance of time, it would be appropriate to speak of the *Yule process* instead of the scheme of autoregression, and the *Slutsky process* instead of the scheme of moving averages. Further the author should like to rename the process of linear regression, calling it instead the *process of moving summation*, a term proposed by A. KOLMOGOROFF [ref. 33]. [p. 3]

[2] SMIRNOFF's test, as well as a related test by KOLMOGOROFF [ref. 30], has the additional advantage that its distribution is independent of the function $F(u)$. [p. 21]

[3] This theorem is known as the *statistical ergodic theorem* of BIRKHOFF-KHINTCHINE. Otherwise expressed, the theorem states that if $\xi_i(t)$ is a fixed realization of the process considered, its average over n time points, that is $\frac{1}{n}\sum_{\nu=0}^{n-1} \xi_i(t-\nu)$, will with probability 1 tend to a limit as $n \to \infty$. Or in yet another reformulation, the statement "with probability 1" means that the limiting average

$$M[\xi] = \lim_{n\to\infty} \sum_{\nu=0}^{n-1} \xi_i(t-\nu)$$

will exist for "almost all" realizations of the process.

The importance of BIRKHOFF-KHINTCHINE's theorem lies in the fact that, whereas the definition of the process $\{\xi(t)\}$ refers to a *universe* of realizations, the theorem makes a statement about an *individual* realization. We note the dualism between the "phase average" $M[\xi]$, which is an average over t for a fixed realization, and the "space average" $E[\xi]$, which is an average over all realizations for a fixed t. Since we are dealing with stationary processes, $E[\xi]$ is the same for all t, whereas $M[\xi]$ will in general vary from one realization to another. For a large class of stationary processes, and in particular for the *ergodic* processes, we have $M[\xi] = E[\xi]$ for almost all realizations, so that the dualism disappears. [p. 35]

[4] Quite generally, let

$$f_t(\xi) = f(\ldots, \xi(t-1), \xi(t), \xi(t+1), \ldots)$$

be a function of random variables $\xi(t)$ that constitute a stationary process $\{\xi(t)\}$. Then

$$f_{t+k}(\xi) = f(\ldots, \xi(t+k-1), \xi(t+k), \xi(t+k+1), \ldots)$$

is obtained from $f_t(\xi)$ by moving all variables ξ simultaneously k steps to the left. We see that $f_{t+k} = f_{t+k}(\xi)$ is a well-defined random variable for all k, and that

$$\{f_t\} = (\ldots, f_{t-1}, f_t, f_{t+1}, \ldots)$$

constitutes a stationary process.

We are now in a position to define the notion of ergodicity: Given a stationary process $\{\xi(t)\}$, we consider an associated process $\{f_t\}$. By the Birkhoff-Khintchine theorem, the phase average

$$M[f] = \lim_{n\to\infty} \frac{1}{n} \sum_{\nu=0}^{n-1} f_{t+\nu}(\xi)$$

will be a well-defined random variable. Process $\{\xi(t)\}$ is called *ergodic* if, for all functions f such that the expectation of f^2 is finite, the phase average $M[f]$ equals the space average $E[f]$ for almost all realizations.

Reference is made to E. Hopf, ref. 25, for the general ergodic theory. As has been noted in section 12, the statistical implications of ergodic theory are of fundamental importance for the applications of stationary processes (see also Appendix notes 14 and 16). [p. 38]

[5] The statements or theorems in question have direct parallels in the theory of statistical scatters or distributions. Reference should here be made to the pioneer work by R. Frisch, ref. 19, where matrix calculus was for the first time systematically employed in statistics. Among the first fruits Frisch reaped thereby were the notion of rank of a statistical scatter, and the theorem that this equals the rank of the corresponding covariance matrix and quadratic form. A parallel theorem on probability distributions has been given by H. Cramér (1937) [Theorem 32 (B)]; for a more detailed treatment, see J. Lukomski, ref. 37, and H. Cramér, ref. 10 (Ch. 22·5]. [p. 44]

[6] Relations equivalent or closely related to (127)–(131) belong to the groundwork in the "generalized harmonic analysis" of N. Wiener (1930). Wiener is primarily concerned with the analysis of an individual time-series, but he extends the analysis to the case of random processes, showing that his methods can be used for a joint simultaneous analysis of the universe of realizations of the process. As applied to stationary processes, Wiener's methods have in the hands of H. Cramér [ref. 9] given further important results. (Cf. also Appendix note 16.) [p. 73]

[7] More precisely, it will be resultless from the viewpoint of forecasting, a main theme of the present work. Over a time interval of any fixed length, the time-series considered can be graduated to any prescribed accuracy by the use of periodogram analysis, but if $\sum|r_k|$ is convergent the graduation will provide no valid forecast outside the interval under analysis. Similarly,

the generalized harmonic analysis of WIENER (see the previous note) is of restricted scope from the viewpoint of forecasting. [p. 74]

[8] In ref. 64 [Ch. 12] the reader will find a detailed exposition of regression analysis from the viewpoint of least-squares approximation in linear spaces. [p. 76]

[9] For a detailed treatment of this and related theorems, see ref. 64 [Theorems 12·3·2–3]. [p. 78]

[10] We have seen that the singular process as defined in section 14 is a straightforward extension of FRISCH's notion of singular (or *collinear*, in FRISCH's terminilogy, ref. 19) distributions in a finite number of dimensions. The extension to processes, however, brings in essentially new features, as seen from our Theorem 2 [p. 45] and from our remarks in section 16 [p. 64]. A further extension is involved in the present definition of singularities of infinite rank. It will be noted that our definition of the singular process is framed with a view to the possibilities of forecasting. In fact, whether the singularity is of finite or infinite rank we see that such a process can be forecasted with any prescribed accuracy on the basis of its past development. More precisely, let $[\ldots, \xi_i(t), \xi_i(t-1), \xi_i(t-2), \ldots]$ denote an arbitrary realization of a singular process $\{\xi(t)\}$; then "almost all" realizations are such that if $\xi_i(t-1)$, $\xi_i(t-2)$, ... are known we can calculate a forecast of $\xi_i(t)$, with a forecast error which has zero expectation and a variance which can be made arbitrarily small.

A. KOLMOGOROFF, refs. 31, 32, has linked up the singularity of a stationary process $\{\xi(t)\}$ with the properties of its spectral function $W(\lambda)$ as defined by formula (116). Specifically, he has shown that a necessary and sufficient condition for $\{\xi(t)\}$ to be singular of finite or infinite rank is that the integral

$$\int_0^\pi |\log W'(\lambda)|\, d\lambda$$

should be infinite.

Regarding the terminology, the author suggested the term *singular process* in view of the relationship with singular matrices and quadratic forms. Better and by now universally accepted is the term *deterministic process* introduced by J. DOOB, ref. 15. Finally we note the alternative term *harmonic process* [see ref. 64] for the process of superposed harmonics dealt with in Theorem 2. Thus a deterministic process is either a harmonic process or a singular process of infinite rank. [p. 81]

[11] This theorem has been deepened and generalized in a brilliant work by A. KOLMOGOROFF, ref. 33. For comment, see Appendix 2 [cf. also ref. 64, Ch. 12·5–7]. [p. 89]

[12] It will be noted that Theorem 8 is *not* restricted to the case when all roots of the characteristic equation of (194) lie within the periphery of the unit circle. [p. 97]

[13] The prediction theory of stationary processes is the subject of important works by KOLMOGOROFF, refs. 31, 32, and WIENER, ref. 60. See also Appendix 2. [p. 103]

[14] In the applications of stationary processes, generally speaking, we are given a time series from which we wish to extract information about the process of which the given series is regarded as a realization. Our device for dealing with this inference problem is to regard the observed serial coefficients $\bar{r}_k$ as large-sample estimates of the theoretical autocorrelation coefficients r_k. The rationale of this device is embodied in the ergodic theorem of BIRKHOFF-KHINTCHINE [Section 12]; in fact, $\bar{r}_k$ and r_k are built up by first and second order moments defined as phase averages and space averages, respectively, and for an ergodic process the two types of average are asymptotically equal. The BIRKHOFF-KHINTCHINE theorem thus is seen to constitute an essential generalization of the large-sample theorems of classical statistics. For further comments on the statistical implications of ergodic theory, see ref. 64 [Chs. 9·4, 10·4, 11·1].

In recent years the theory of time-series inference has made rapid progress. Briefly stated, we may distinguish two stages in the development. In the first stage the typical problem is to establish the distribution of $\bar{r}_k$, or of any other parameter estimate, on the basis of specific assumptions about the generating process. In the second stage the inference is not restricted to a single parameter, but refers to the entire structure of the generating process. For the treatment of the deep-lying problems of the second type, a new line of approach has recently been opened up by P. WHITTLE, refs. 54–59. Constructed on a least-squares basis, Whittle's methods have certain optimal properties as large-sample procedures, and they lead to a uniform treatment of large classes of stationary processes, including the processes of moving summation. WHITTLE gives an expository survey of his methods in Appendix 2, in which also some fresh results are incorporated. [p. 109]

[15] Another case of interest is $p = q$, giving $b_k = r_k = (k+1)p^k$. (Cf. also p. 145.) [p. 113]

[16] In Appendix notes 3, 4 and 14 we have touched upon the fact that our theoretical analysis in Chapters II–III and in particular the notion of theoretical correlogram refers to a stochastic process, i.e. to a universe of hypothetical time-series (realizations), whereas in the applications we are usually concerned with a single time-series, and in particular the empirical correlogram is usually based on a single series. For ergodic processes the dualism between universe and single realization is not essential, for we know that an individual realization will then suffice to give full information about the whole universe.

The dualism between universe and realization has been discussed from statistical viewpoints in a later paper, ref. 62 (see also ref. 64, Ch. 12·5–7). On the assumption that the time-series under analysis is given from the infinite past up to a fixed time-point t_0, it is shown that the series can be subjected to a decomposition analogous to that in Theorem 7. The decomposition applies, in particular, to an individual realization of a stationary process. For ergodic processes, this decomposition will with probability 1 coincide with that given by Theorem 7. For non-ergodic processes, on the other hand, the two decompositions will in general differ, the one based on an individual realization giving components the structure of which may vary from one realization to another. From the viewpoint of the applications it is important

to note, however, that the information extracted from the past of an individual realization can always be employed for forecasting the future development of the same realization. [p. 147]

[17] Stochastic processes in several dimensions or variables is a wide field of research, which in recent years has been explored in several directions; reference is made to the fundamental works of H. CRAMÉR, ref. 8, V. ZASUHIN, ref. 65, H. B. MANN and A. WALD, ref. 39, and P. WHITTLE, ref. 59. The scheme (320) falls under the heading of the *recursive process*, a type of multidimensional process which merits particular attention because of its general scope in the applications. The pioneer on this line is TINBERGEN (1937), who has followed up the approach in further important investigations [refs. 47, 48]. Later, the theory and the application problems of the recursive process have been studied in some detail by the author [see R. BENTZEL and H. WOLD, ref. 6, and refs. 62, 63; the results have been summed up in ref. 64 (Chs. 3·2 and 12·7). See also refs. 5, 49, 50]. [p. 189]

Appendix 2

by P. WHITTLE

Some recent contributions to the theory of stationary processes

In view of the necessarily short compass of this article it has been deemed better to include as a rule only such work as fits into a uniform treatment, rather than to attempt a faithful review of later years' literature. Further, while there are aspects of neighbouring fields, such as the more general theory of stochastic processes, which are far from irrelevant to the present discussion, they must regretfully be left to such treatments as can do them justice.

The two theories whose application has most helped time series analysis during the last fifteen years are those of *spectral analysis* and *statistical inference.*

Referring to equation (115) of Professor WOLD's work, spectral theory may be described as the study of the function $W(x)$, together with its exploitation to classify the different types of process, and to deduce the existence of stochastic relations among the process variates. This last aspect, the deduction of stochastic relations, includes what is generally known as *prediction theory.* It is common enough that we assume some relation to hold between the variates (e.g., an autoregression) and then calculate the spectrum, etc. Prediction theory goes rather in the opposite direction, and seeks to establish such relations, given the spectrum. WOLD's decomposition (Theorem 7) was the first proof that such a relation exists for a general class of cases. Of course, the ultimate aim of most time series analyses is to obtain a forecast, so that the idea of prediction is latent in the whole subject.

It is useful to keep in mind the intuitive meaning of $W(x)$. If the series is regarded as describing some electrical or mechanical oscillation, then $W(x)$ is the theoretical distribution function of the oscillation power among the different frequency components. This distribution gives surprisingly de-

tailed information on the mechanism of generation of the series — hence its importance.

We use the word "inference" in its usual sense: the endeavour to form a general hypothesis from observed facts; in this case, the setting up of a hypothetical model to explain as nearly as possible the generation of a given series. More specifically, in statistical inference one sets up criteria measuring the degree of agreement of hypothesis and observation (tests of fit), or discriminating between different hypotheses (discriminatory tests), and one endeavours to estimate numerical quantities involved in the hypotheses.

The theory of inference is a strange mixture of mathematical and non-mathematical considerations. It is mathematical insofar as that criteria such as those mentioned above may be constructed, even although they may often have a flavour of arbitrariness. But when we ask, which hypotheses shall be tested? what a priori weightings shall be given them? then we enter a realm which is perhaps inaccessible to exact method, and where full certainty can never be reached. The pursual of these questions would be profitless here, however, and we shall not continue it until we are in a position to do so more exactly.

Chapter 1. Spectral theory

1·1. *Spectral representation*

It is convenient to modify slightly the notation of WOLD's § 17, so that equations (114), (115) become

$$\varrho(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{iu\omega} \, dV(\omega) \tag{1}$$

$$\varrho_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\omega} \, dW(\omega) \tag{2}$$

the spectral representations of the autocovariance function of a continuous and of a discrete process, respectively. As in that section, the inverse relations hold:

$$V(\omega) = \int_{-\infty}^{+\infty} \frac{e^{iu\omega}}{iu} \varrho(u) \, du \tag{3}$$

$$W(\omega) = \sum_{-\infty}^{+\infty} \frac{e^{ik\omega}}{ik} \varrho_k . \tag{4}$$

It is already evident that the theories of a continuous and a discrete process are very similar. The former is actually the more general, but if we nevertheless restrict ourselves to the discrete case, we shall avoid mathematical difficulties which are most often irrelevant to the physical problem, and still have a treatment general enough for the overwhelming majority of practical cases.

Now, from the Fourier representation of the autocovariance (2) CRAMÉR has deduced a similar representation of the process variate

$$x_t = \int_{-\pi}^{\pi} e^{it\omega}\, d\, y(\omega) \tag{5}$$

[see ref. 9]. The equality (5) must be understood to hold as a limit in mean square. $\{y(\omega)\}$ is also a stochastic process, a so-called *process of noncorrelated increments*. That is

$$E[\{y(\omega_1) - y(\omega_2)\}\,\overline{\{y(\omega_3) - y(\omega_4)\}}] = 0 \quad \text{if} \quad \omega_1 \gtrless \omega_2 \gtrless \omega_3 \gtrless \omega_4. \tag{6}$$

This has as consequence that, at least when ω_1 and ω_2 are continuity points of $W(\omega)$,

$$W(\omega_1) - W(\omega_2) = 2\pi\, E\,[\,|\,y(\omega_1) - y(\omega_2)\,|^2] \qquad (\omega_1 \gtrless \omega_2). \tag{7}$$

That is, the increment of the spectrum in a certain interval is proportional to the mean square of the corresponding y increment in that interval. Should the spectrum there not increase at all, then the y increment in any part of that interval is equal to zero, in mean square.

1·2. *Decomposition of the variate*

CRAMÉR's representation (1·1·5) may be regarded as the limiting case of a graduation of the process in terms of harmonic functions of time; more precisely, a graduation over a finite interval $(-T, +T)$ is obtained by Fourier analysis, and then the interval is extended by letting $T \to \infty$. WOLD's representation (Theorem 7) is on the other hand derived from a representation of the process in terms of its own past. Thus, briefly stated, CRAMÉR proceeds in terms of graduation, WOLD in terms of forecasting. In view of the apparent divergence of these two approaches, it is of interest to note that they lead to results which are very nearly related, although it is no trivial matter to establish precisely what is equivalent and what is different. We shall briefly review some work of CRAMÉR, KOLMOGOROFF and WIENER [refs. 9, 31–33, 60; see also WIENER (1930)] in the field.

The values of the derivative of the spectrum between 0 and 2π constitute in effect the assembly of eigenvalues of the infinite covariance matrix of

the process $(\ldots x_{t-1}, x_t, x_{t+1} \ldots)$, [see ref. 45 p. 201, ref. 54 p. 36]. Since a finite set of statistical variates identically obeys some linear relation if its covariance matrix possesses one or more zero eigenvalues, we may expect the general result that the $\{x_t\}$ process will be singular (deterministic) in the sense of section 14 if this derivative be anywhere zero, i.e. if the spectrum be anywhere nondecreasing. We are dealing with an infinite number of variates, however, and a continuous assembly of eigenvalues, so that the actual results are in general of a more sophisticated nature.

Cramér [ref. 9] proceeded from the classical decomposition of a monotone increasing function

$$W(\omega) = W_1(\omega) + W_2(\omega) + W_3(\omega). \tag{1}$$

Here $W_1(\omega)$, $W_2(\omega)$ and $W_3(\omega)$ are nondecreasing, and are, respectively, an absolutely continuous function, a step function, and a so-called singular function, continuous, but constant almost everywhere. Define now the three processes $y_1(\omega)$, $y_2(\omega)$, $y_3(\omega)$ such that $y_j(\omega)$ is constant except at the points of increase of $W_j(\omega)$, $(j = 2, 3)$, where it has an increment equal to that of $y(\omega)$; $y_1(\omega) = y(\omega) - y_2(\omega) - y_3(\omega)$. Then the process variate may be decomposed into

$$x_t = x_1(t) + x_2(t) + x_3(t) = \sum_{j=1}^{3} \int_{-\pi}^{\pi} e^{it\omega}\, d\, y_j(\omega). \tag{2}$$

It follows from (1·1·6) that the three components are uncorrelated, and that their spectra are $W_1(\omega)$, $W_2(\omega)$ and $W_3(\omega)$ respectively.

If $y_2(\omega)$ has a finite number of points of increase, $x_2(t)$ is the sum of a finite number of harmonic terms, and so, as Wold reasons, constitutes a so-called singular or nondeterministic component. In general, however, the series of discontinuities of $y_2(\omega)$ will be denumerable, but it may be shown [ref. 31; cf. also ref. 44] that $x_2(t)$ is still deterministic. The component $x_3(t)$ is not so easily characterised, but may also be shown to be deterministic, as the largely constant nature of its spectrum would lead us to expect.

We see thus that both $x_2(t)$ and $x_3(t)$ must be relegated to $\psi(t)$, the deterministic component of Wold's representation. Of the remaining component, $x_1(t)$, it may be said that it is *either* purely nondeterministic *or* deterministic (nonsingular or singular, in Wold's terminology) according as the condition

$$\int_{-\pi}^{\pi} \left| \log W_1'(\omega) \right| d\,\omega < \infty \tag{3}$$

is fulfilled or not (this is Kolmogoroff's result, refs. 31, 32). To see this, we define

$$Y^*(\omega) = \int^{\omega} \frac{d y_1(\omega)}{\sqrt{W_1'(\omega)}} \tag{4}$$

and

$$\eta_t^* = \int_{-\pi}^{\pi} e^{it\omega}\, d\,Y^*(\omega). \tag{5}$$

It is readily verified that the $\eta^*{}_t$ variates are mutually uncorrelated and have unit variance, and that

$$x_{1\,t} = \int_{-\pi}^{\pi} e^{it\omega} \sqrt{W_1'(\omega)}\, d\,Y^*(\omega) = \sum_{s=-\infty}^{+\infty} d_s\, \eta_{t-s}^* \tag{6}$$

if $\sqrt{W_1'(\omega)}$ is expanded in a formal Fourier series $\sum_{-\infty}^{\infty} d_s e^{-i\omega s}$. That is, a process with absolutely continuous spectrum may be represented as a moving average of a sequence of uncorrelated variates. Conversely, a process thus representable must have absolutely continuous spectrum [ref. 33; see also ref. 16].

If condition (3) is fulfilled, then the representation (6) may be specialised to that of a *one-sided* moving average, where s assumes values from 0 to $+\infty$ so that $x_1(t)$ is purely nondeterministic (it is in this case the component $\zeta(t)$ of WOLD's decomposition). For then $\log W_1'(\omega)$ may be expanded in a symmetric Fourier series

$$\log W_1'(\omega) = \sum_{-\infty}^{-\infty} c_s\, e^{-i\omega s} \tag{7}$$

and, defining

$$\theta(\omega) = e^{\sum_1^{\infty} c_s\, e^{-i\omega s}} = 1 + b_1 e^{-i\omega} + b_2 e^{-2i\omega} + \cdots \tag{8}$$

we have

$$W_1'(\omega) = e^{c_0} \theta(\omega)\, \theta(-\omega). \tag{9}$$

That is, we have succeeded in breaking $W_1'(\omega)$ up into two conjugate factors, one of which may be expanded in nonpositive powers of $e^{i\omega}$. Defining now

$$Y(\omega) = \int_{-\pi}^{\omega} \frac{d y_1(\omega)}{\theta(\omega)} \tag{10}$$

$$\eta_t = \int_{-\pi}^{\pi} e^{i\omega t}\, d\,Y(\omega) \tag{11}$$

we find as before that the η_t's are uncorrelated (although they have now variance e^{c_0}, a consequence of the fact that the leading coefficient in (8) was chosen as unity) and that

$$x_1(t) = \int_{-\pi}^{\pi} e^{it\omega}\,\theta(\omega)\,dY(\omega) = \sum_0^\infty b_s\,\eta_{t-s}. \tag{12}$$

This is the required representation.

If, on the other hand, condition (3) should be violated, then $x_1(t)$ is actually deterministic, despite the apparent nondeterminacy of representation (6). This may be roughly seen by representing the process with spectrum $c\omega + W_1(\omega)$ (c a constant) as an infinite autoregression. Upon letting c tend to zero, it is found that the variance of the error term also tends to zero, so that $x_1(t)$ is determined by past values, although the autoregression may not exist as such.

1·3. *Linear operations*

We saw in the previous section that the general purely non-deterministic process could be represented as a one-sided moving average

$$x_t = \sum_0^\infty b_s\,\eta_{t-s}. \tag{1}$$

Now, the forming of this moving average may be regarded as an operation upon the η process, an operation which is summed up in the function $\theta(\omega)$. For instance, the operation of a finite moving average

$$x_t = \sum_0^p b_s\,\eta_{t-s} \tag{2}$$

corresponds to $\theta(\omega) = b_0 + b_1 e^{-i\omega} + \cdots + b_p e^{-ip\omega}$, while the autoregression

$$\begin{aligned} x_t - \varrho x_{t-1} &= \eta_t \\ \text{or} \qquad x_t &= \eta_t + \varrho\,\eta_{t-1} + \varrho^2\,\eta_{t-2} + \cdots \end{aligned} \tag{3}$$

corresponds to $\theta(\omega) = 1 + \varrho e^{-i\omega} + \varrho^2 e^{-2i\omega} + \cdots = (1 - \varrho e^{-i\omega})^{-1}$.

This is one of the great virtues of spectral analysis, that a linear operation upon a process is equivalent to a certain function. Thus, the calculus of such operations may be interpreted in terms of a much more familiar calculus of functions.

Suppose that we represent the operation by L, i.e.

$$L(\eta_t) = b_0\,\eta_t + b_1\eta_{t-1} + b_2\eta_{t-2} + \cdots \tag{4}$$

and write the equivalence as $L \doteq \theta(\omega)$. Then the following rules are fairly simply deduced from equation (11) of the preceding section:

$$L_1 + L_2 = L_2 + L_1 \doteqdot \theta_1(\omega) + \theta_2(\omega) \tag{5}$$

$$L_1 L_2 = L_2 L_1 \doteqdot \theta_1(\omega)\, \theta_2(\omega) \tag{6}$$

$$L^{-1} \doteqdot [\theta(\omega)]^{-1}. \tag{7}$$

Equation (7) holds as it stands only if L^{-1} is also of type (4), i.e. if $[\theta(\omega)]^{-1}$ has also a Taylor expansion in $e^{-i\omega}$.

The following may serve as an example of the application of these rules. Consider the autoregressive scheme

$$x_t + a_1 x_{t-1} + \cdots + a_p x_{t-p} = \eta_t. \tag{8}$$

This equation may be written

$$L^{-1} x_t = \eta_t. \tag{9}$$

Comparing (8), (9) we see that

$$L^{-1} \doteqdot 1 + a_1 e^{-i\omega} + \cdots + a_p e^{-ip\omega}. \tag{10}$$

Thus, by (7)

$$L \doteqdot (1 + a_1 e^{-i\omega} + \cdots + a_p e^{-ip\omega})^{-1} \tag{11}$$

which gives $\theta(\omega)$ for the scheme (8). Since all singularities of $\theta(\omega)$ are in the upper half-plane, the application of (7) is valid.

We see from (6) that the spectral function [i.e., the differentiated spectrum] of the process obtained by operating upon the η process with the operators L_1, L_2 in turn, is

$$|\theta_1(\omega)\theta_2(\omega)|^2 \sigma^2(\eta) = |\theta_1(\omega)|^2 |\theta_2(\omega)|^2 \sigma^2(\eta) \tag{12}$$

i.e., the product of the individual spectral functions of the processes $\{L_1(\eta_t)\}$, $\{L_2(\eta_t)\}$, apart from the factor $\sigma^2(\eta)$.

1·4. *Rational spectral functions*

When dealing with a discrete process, it is convenient to make the variable transformation $z = e^{i\omega}$. Thus, for a purely nondeterministic process

$$F(\omega) = W'(\omega) = \sum_{-\infty}^{+\infty} \varrho_s z^s = \Lambda(z) \tag{1}$$

say, and

$$\varrho_s = \frac{1}{2\pi i} \int_c z^{s-1} \Lambda(z)\, dz \tag{2}$$

where c is the unit circle in the z plane, positively described.

A class of processes of particular interest is that for which $\Lambda(z)$ is rational in z. From the fact that $\Lambda(z)$ is rational and is real on the unit circle, we

deduce by the Schwarz reflection principle that if it has a zero (pole) at α, then it also possesses a zero (pole) at $\bar{\alpha}^{-1}$. Hence $\Lambda(z)$ may be written

$$\Lambda(z) = K \frac{\Pi (z-\alpha)(\bar{\alpha} z - 1)}{\Pi (z-\beta)(\bar{\beta} z - 1)} = K \frac{|\Pi (z-\alpha)|^2}{|\Pi (z-\beta)|^2} = K \left| \frac{B(z)}{A(z)} \right|^2 \tag{3}$$

where the polynomial $B(z)$ may be chosen so that all its roots lie within or on the unit circle, and $A(z)$ so that all its roots lie within the unit circle, (obviously $A(z)$ cannot have a root *on* the unit circle, for this would cause $W(\omega)$ to have a singularity in $(0, 2\pi)$, against hypothesis). For a real process, which is the only kind we consider, $\Lambda(z)$ is an even function of ω, i.e. $\Lambda(e^{-i\omega}) = \Lambda(e^{-i\omega})$, so that we may suppose the polynomials $A(z)$, $B(z)$ to have real coefficients. We may thus write, for $|z| = 1$,

$$\Lambda(z) = K \frac{B(z)\overline{B(z)}}{A(z)\overline{A(z)}} = K \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})}. \tag{4}$$

Referring to section 1·2, we see that a possible choice of $\theta(\omega)$ is

$$\theta(\omega) = z^{a-b} \frac{B(z)}{A(z)} \tag{5}$$

which is readily seen to correspond to the operation

$$x_t + a_1 x_{t-1} + \cdots + a_a x_{t-a} = \eta_t + b_1 \eta_{t-1} + \ldots b_b \eta_{t-b}. \tag{6}$$

Equation (6) illustrates the structure of a process having spectral function rational in z. This class embraces the two important schemes for a discrete stationary variate distinguished by WOLD: the *moving average* scheme and the *autoregressive* scheme, given by setting b, a respectively equal to zero.

These schemes are usually defined directly from the stochastic difference equation (6) with the η's defined as an independent sequence. The approach here [due to DOOB, ref. 16] has been in the reverse direction, so as to speak, but is instructive insofar as it shows that any spectral function rational in z may be regarded as a consequence of an operation of type (6). Furthermore, so far as spectral and autocovariance theory is concerned, the variates η_t need only be assumed uncorrelated.

Without doubt, the importance of the rational spectral function depends to a large extent upon its theoretical tractability, and the attractive simplicity of (6). In technological applications the restriction of rationality is a very natural one, however, since any electrical or mechanical filter constructed entirely of "linear" components will perform an operation of type

(6) upon the input. (Actually, the quotient of input and output spectral functions will be rational in ω, but if we take discrete equidistant observations it will then appear to be rational in $e^{i\omega}$.)

1·5. *The intrinsic variance*

We see from the decomposition

$$x_t = \psi_t + \eta_t + b_1\eta_{t-1} + b_2\eta_{t-2} + \cdots \tag{1}$$

that η_t is the random element which has entered the process in the time interval $(t-1,\ t)$, and is consequently that part of x_t which cannot be deduced from a knowledge of $x_{t-1},\ x_{t-2}\ldots$ (which in turn implies a knowledge of $\eta_{t-1},\ \eta_{t-2}\ldots$). It is appropriate, then, that $v = \sigma^2(\eta)$ be called the *prediction variance* or *intrinsic variance* of the process.

Now, we saw from section 1·2 that the η variates have variance e^{c_0}, where c_0 is the absolute term in the Fourier expansion of $\log W_1'(\omega)$. Thus

$$v = e^{\frac{1}{2\pi}\int\limits_0^{2\pi} \log W_1'(\omega)\, d\omega} \tag{2}$$

That is, the prediction variance is equal to the continuous geometric mean of $W_1'(\omega)$. This equality was proved by KOLMOGOROFF [refs. 31, 32] in 1941, although SZEGÖ had earlier obtained virtually the same result [refs. 45, 46] in a non-statistical paper.

We shall define the *normalised spectral function*, $G(\omega)$ or $M(z)$, as

$$G(\omega) = M(z) = \frac{W_1'(\omega)}{v}. \tag{3}$$

It is obvious that $G(\omega) = |\theta(\omega)|^2$, and furthermore that

$$\int\limits_0^{2\pi} \log G(\omega)\, d\omega = 0. \tag{4}$$

$G(\omega)$ is that part of the spectral function which refers to the linear operation L [see eq. (1·2·12)] and so is of primary importance. However, when we come to the inductive side of the problem we shall find that v is of considerable importance too, since the touchstone of a hypothetical model is that it lead to as small a prediction variance as possible.

1·6. *Prediction*

WOLD's decomposition has the nature of an existence theorem, but as soon as the deterministic component and the coefficients of the moving average are specified it becomes a *linear prediction formula*.

The subject of prediction has been much studied under recent years, the two outstanding contributions being undoubtedly the parallel works of KOLMOGOROFF [refs. 31, 32, 33] and WIENER [ref. 60]. WIENER's work is the less general, but the more readily applicable. The first problem he sets himself, and the only one we shall consider here, is the extrapolation of a stationary purely nondeterministic series. We shall record his results for the discrete case.

If the extrapolation is one of α steps, then the criterion that the difference between the forecast

$$X_{t+\alpha} = \sum_{0}^{\infty} K_\nu x_{t-\nu} \tag{1}$$

and reality, $x_{t+\alpha}$, have minimum variance over all realisations leads to the equation

$$\varrho_{\tau+\alpha} = \sum_{0}^{\infty} K_\nu \varrho_{\tau-\nu} \qquad (\tau = 0, 1, 2 \ldots) \tag{2}$$

which will be recognised as a limit form of the usual estimation equations for the coefficients of an autoregressive scheme [see WOLD's equations (221), (312)].

An explicit solution is obtained for K_ν in the following manner. We recall the function $\theta(\omega)$ of section 1·2 for which

$$W'(\omega) = e^{c_0} |\theta(\omega)|^2 = e^{c_0} |1 + b_1 e^{-i\omega} + b_2 e^{-2i\omega} + \cdots|^2. \tag{3}$$

Let us also define

$$k(\omega) = \sum_{\nu} K_\nu e^{-i\nu\omega} \tag{4}$$

so that

$$K_\nu = \frac{1}{2\pi} \int_{0}^{2\pi} e^{i\nu\omega} k(\omega)\, d\omega. \qquad (\nu = 0, 1, 2 \ldots) \tag{5}$$

WIENER shows then that

$$k(\omega) = \frac{1}{2\pi\,\theta(\omega)} \sum_{0}^{\infty} e^{-i\nu\omega} \int_{0}^{2\pi} \theta(u)\, e^{i(\nu+\alpha)u}\, du \tag{6}$$

which may alternatively be written

$$k(\omega) = \frac{\sum_{s=0}^{\infty} b_{s+\alpha} e^{-is\omega}}{\sum_{s=0}^{\infty} b_s e^{-is\omega}}. \tag{7}$$

The prediction error variance is $e^{c_0} \sum_{0}^{\alpha-1} |b_\nu|^2$.

The b_ν coefficients of (3) are in fact the coefficients of a moving average prediction formula [see eq. (1·2·12)]. In equations (6), (7) this moving average formula is transformed to an autoregressive prediction formula. The autoregressive form is obviously preferable, since the forecast is then expressed directly in terms of observed quantities.

When $\alpha = 1$ the prediction error variance is equal to e^{c_0}, in accordance with the result of the previous section.

It should perhaps be remarked that the numerical calculation of formulae such as (1) gives merely an optimum graduation of the series, and cannot in itself give any new insight into the generation mechanism of the process. Nor, indeed, can any method of analysis which proceeds purely by rote.

Chapter 2. Inference

2·1. *Inference in time series*

We shall now pass over to the inductive side of the problem: given an empirical series, what is the process which generated it? This is the question which is asked first and last; we investigate particular models only that we may better provide an answer.

The subject of time series estimation and testing tends to be obscured by haze, and this must be due quite simply to the difficulty of evaluating the expectations and distributions which arise, since there is nothing in the actual inference problems which is special just for time series. We shall therefore begin with a discussion of the distribution of certain regularly encountered sample functions. We shall see that the estimation equations and test statistics are completely analogous to those occurring in the study of independent variates, even if they do not hold so exactly. It will be further remarked that the familiar least-squares criterion is well in evidence. The reason is partly that both spectral theory and least squares theory involve only the second moments of the observations, and therefore complement one another very naturally.

2·2. *The relation between observed and residual moments*

There are very few statistics in time series whose distributions may be evaluated exactly, and approximations are the rule rather than the exception. One of the stumbling blocks in the way of exact analysis is the "end effect" of a finite series, which must usually be neglected, the justification being that it is sensible only for a short series.

In this way, a number of useful results connecting the autocovariances of the observed and residual variates may be derived directly from the moving average representation of a purely nondeterministic series

$$x_t = \eta_t + b_1 \eta_{t-1} + b_2 \eta_{t-2} + \cdots \tag{1}$$

or the corresponding autoregressive representation (when it exists!)

$$x_t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots = \eta_t \tag{2}$$

if we remember that

$$G(\omega) = \frac{F(\omega)}{v} = |1 + b_1 e^{-i\omega} + b_2 e^{-2i\omega} + \cdots|^2 = |1 + a_1 e^{-i\omega} + a_2 e^{-2i\omega} + \cdots|^{-2}. \tag{3}$$

Thus, suppose that the observed series, $x_1, x_2 \ldots x_n$, has autocovariances

$$C_s = \frac{1}{n-s} \sum_{1}^{n-s} x_t x_{t+s} \tag{4}$$

and periodogram

$$f(\omega) = \frac{1}{n} [(\Sigma x_t \sin \omega t)^2 + (\Sigma x_t \cos \omega t)^2] \approx \sum_{-n}^{+n} C_s e^{i\omega s} \tag{5}$$

while the corresponding quantities for the residual series $\eta_1, \eta_2, \ldots \eta_n$ are $C_s^{(\eta)}$ and $f^{(\eta)}(\omega)$. We find then readily from (2) and (3) that

$$f(\omega) \approx G(\omega) f^{(\eta)}(\omega) \tag{6}$$

and

$$C_s^{(\eta)} \approx \sum_{\nu} \gamma_\nu C_{\nu+s} \tag{7}$$

where $[G(\omega)]^{-1} = \sum \gamma_s e^{i\omega s}$. Equation (7) can be rewritten

$$C_s^{(\eta)} \approx \frac{1}{2\pi} \int_0^{2\pi} \frac{f(\omega)}{G(\omega)} e^{i\omega s} \, d\omega. \tag{8}$$

We shall find equations (7) and (8) especially useful later for the case $s = 0$, when they obviously give us the *residual sum of squares* of the observed series, expressed in terms of the correlogram and periodogram respectively.

A condition that (7) and (8) hold is that $[G(\omega)]^{-1}$ shall possess a Fourier expansion, which is also the condition that representation (1) be transformable to (2). This condition is seldom irksome in practice, although it means that we cannot deal with processes such as

$$x_t = \eta_t - \eta_{t-1}$$

for which $G(0) = 0$, and for which (2) takes the nonconvergent form

$$\eta_t = x_t + x_{t-1} + x_{t-2} + \cdots \qquad \text{[see however ref. 61]}$$

2·3. *Expressions for the sampling moments*

Most of the sample functions of interest in time series analysis are quadratic functions of the observations, whose distribution properties depend to a large extent upon the corresponding matrix of the quadratic form.

The results of the last section can be derived in a more sophisticated manner by matrix methods [see ref. 54]. For example, eq. (2·2·7) can be written, for $s = 0$,

$$U = \eta' \eta \approx x' [M(W)]^{-1} x \tag{1}$$

where x, η are the observation and residual vectors, $M(z) = G(\omega)$ [see section 1·5], and W is the circulant matrix

$$W = \begin{bmatrix} \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}. \tag{2}$$

In (1), C_s has been approximated by the so-called "circular autocovariance"

$$C_s = \frac{1}{n} x' W^s x. \tag{3}$$

This modification of the usual definition was first introduced by HOTELLING [see ref. 1], and is in effect an elegant method of neglecting the end effect. We shall apply it to calculate the cumulants of a linear function of the autocovariances, under the assumption that the residual variates are distributed normally.

Consider the linear function of the autocovariances

$$\xi = x' Q(W) x \tag{4}$$

where it is assumed that $Q(e^{i\omega})$ may be expanded in a symmetric Fourier series (symmetry can always be arranged, since $x' W^s x = x' W^{-s} x$). Then, if $E(XX') = V$, the characteristic function of ξ is the determinant

$$\Phi(\theta) = |I - 2i\theta V Q(W)|^{-\frac{1}{2}} \approx |I - 2i\theta \Lambda(W) Q(W)|^{-\frac{1}{2}} \tag{5}$$

since $V \approx vM(W) = \Lambda(W)$, by (1). Now, a function of W, say $h(W)$, has in general latent roots $h(\varepsilon^j)$, $(j = 1, 2 \ldots n)$, where $\varepsilon = e^{\frac{2\pi i}{n}}$, [see ref. 54]. Thus

$$\Phi(\theta) \approx \prod_{j=1}^{n} [1 - 2i\theta \Lambda(\varepsilon^j) Q(\varepsilon^j)]^{-\frac{1}{2}} \tag{6}$$

so that

$$\psi(\theta) = \log \Phi(\theta) \approx -\tfrac{1}{2}\sum_{1}^{n} \log [1 - 2i\theta \Lambda(\varepsilon^j) Q(\varepsilon^j)] \tag{7}$$

$$\approx -\frac{n}{4\pi i} \int\limits_{|z|=1} \log [1 - 2i\theta \Lambda(z) Q(z)] \frac{dz}{z} \tag{8}$$

if we approximate the sum by an integral. If $Q(z) = P(\omega)$, then (8) may be written

$$\psi(\theta) \approx -\frac{n}{4\pi} \int\limits_{0}^{2\pi} \log [1 - 2i\theta F(\omega) P(\omega)]\, d\omega \tag{9}$$

and, expanding this expression in θ, we obtain the following formula for the cumulants of ξ:

$$k_j \approx \frac{2^{j-1}(j-1)!\, n}{2\pi} \int\limits_{0}^{2\pi} [F(\omega) P(\omega)]^j\, d\omega. \tag{10}$$

We shall find constant use for these two equations. Note that n enters only as a simple scale factor, i.e. $\Phi(\theta)$ is approximately of the form $[\alpha(\theta)]^n$, and we have reduced the consideration of an autocorrelated series to that of an independent series.

In the generalisation of (9), (10) to the multivariate case $\theta P(\omega)$ will be simply replaced by $\Sigma_j \theta_j P_j(\omega)$.

2·4. *Miscellaneous results*

There are a number of miscellaneous results which should be mentioned, although they do not fit into any uniform treatment.

Koopmans [ref. 34] proved eq. (2·3·9) for the special case $F(\omega) = 1$, $P(\omega) = \cos \omega$, when investigating the distribution of the first autocorrelation coefficient of a random series. Continuing along this line it was shown [refs. 12, 43] that the frequency function of the coefficient is

$$f(r) \approx \text{const.}\, (1 - r^2)^{\frac{n-1}{2}}. \tag{1}$$

This result is valid for any of the earlier autocorrelation coefficients of a random series, since expression (2·3·8) is unaltered in value if z is replaced by z^p, p integral.

Madow [ref. 38] has extended (1) to the case where the variates follow a first order autoregression with neighbour correlation ϱ:

$$f(r) \approx \text{const.} \left[\frac{1 - r^2}{1 - 2\varrho r + \varrho^2}\right]^{\frac{n-1}{2}}. \tag{2}$$

QUENOUILLE [ref. 41] has generalised (2) to give the simultaneous distribution of $r_1, r_2 \dots r_k$, when the autoregression is of order k. Analogous expressions have not yet been obtained for any more general case, to the author's knowledge. Of course, the frequency function for any particular statistic can in general be approximated by a type A series with the help of the expressions for the cumulants given by our equation (2·3·10), and there are other possibilities, all yet relatively untried (e.g., the type C representation [ref. 24] and the method of mixtures [ref. 42]).

A pair of interesting relations are BARTLETT's formulae [ref. 2] for the covariances of the empirical autocovariances and autocorrelation coefficients

$$\operatorname{cov}(C_s, C_t) \approx \frac{v^2}{n}[X_{s+t} + X_{s-t}] + K \tag{3}$$

$$\operatorname{cov}(r_s, r_t) \approx \frac{1}{n}[X_{s+t} + X_{s-t} + 2\,\varrho_s\,\varrho_t\,X_0 - 2\,\varrho_t\,X_s - 2\,\varrho_s\,X_t]. \tag{4}$$

Here K is a function of the fourth cumulant of the residual variate η, $\varrho_s = E(x_t x_{t+s})/E(x_t^2)$, and $X_s = \sum \varrho_v\,\varrho_{v+s}$. If we set $j = 2$ in formula (2·3·10) we clearly obtain a relation equivalent to (3) for the normal case, when $K = 0$. Formula (4) is interesting in that it does not involve any of the moments of the residual variate. We can very simply show that this property is a general one for the ratios of quadratic forms occurring in the analysis of stationary series.

Consider the quadratic forms in the residuals

$$a = \eta' A \eta \qquad b = \eta' B \eta.$$

If η_t has cumulants, 0, $k_2, k_3, k_4 \dots$, then we readily find that

$$\bar{a} = E(a) = k_2 \operatorname{tr} A$$
$$\operatorname{cov}(a, b) = k_4 \Sigma_i\, a_{ii}\, b_{ii} + 2k_2^2 \operatorname{tr} AB.$$

Here we have assumed that the matrices A, B are symmetric and have elements a_{ij}, b_{ij}, so that $\operatorname{tr} A = \Sigma\, a_{ii}$.

Consider now the ratio $r = a/b$. We find

$$E(r) = \frac{\bar{a}}{\bar{b}}(1 + 0(n^{-\frac{1}{2}})) = \frac{\operatorname{tr} A}{\operatorname{tr} B}(1 + 0(n^{-\frac{1}{2}})) \tag{5}$$

and

$$\operatorname{var}(r) = \left[\left(\frac{1}{\bar{b}}\right)^2 \operatorname{var}(a) + \left(\frac{\bar{a}}{\bar{b}^2}\right)^2 \operatorname{var}(b) - 2\left(\frac{\bar{a}}{\bar{b}^3}\right) \operatorname{cov}(a, b)\right](1 + 0(n^{-\frac{1}{2}})) =$$

$$= \frac{(1 + 0(n^{-\frac{1}{2}}))}{\bar{b}^4}\begin{bmatrix} k_4[\bar{b}^2 \Sigma\, a_{ii}^2 + \bar{a}^2 \Sigma\, b_{ii}^2 - 2\,\bar{a}\,\bar{b}\,\Sigma\, a_{ii} b_{ii}] + \\ + 2\,k_2^2[\bar{b}^2 \operatorname{tr} A^2 + \bar{a}^2 \operatorname{tr} B^2 - 2\,\bar{a}\,\bar{b} \operatorname{tr} AB] \end{bmatrix}. \tag{6}$$

Now, the coefficient of k_4 in (6) can vanish under a variety of conditions, and in particular it will vanish when a_{ii} and b_{ii} are independent of i. This will approximately be the case for all quadratic forms occurring in the analysis of a stationary time series, since these will have matrices which are approximately Laurent matrices [see equation (2·3·1) for example] so that the elements of the principal diagonal are approximately constant. That is, all quotients which we shall expect to have to deal with have means and variances [and covariances, by the same proof] which are asymptotically independent of the residual variate's distribution function. This is a result of some importance, since our knowledge of the η distribution in any practical case is usually of the scantiest.

2·5. *Estimation, nondeterministic case*

We shall adopt the least square estimation criterion, so that the estimating relation is obtained by minimising the residual sum of squares as given by (2·2·8). That is, if $G(\omega)$ and v have least squares estimates $\hat{G}(\omega)$ and $\hat{v}$, then

$$\frac{1}{2\pi}\int_0^{2\alpha} \frac{f(\omega)}{\hat{G}(\omega)}\,d\omega = \text{min.} = \hat{v}. \tag{1}$$

The minimisation is performed with respect to the unknown parameters of $G(\omega)$, which we shall denote by $\theta_1, \theta_2 \ldots \theta_p$. However, the minimisation (1) is not a free one, since $G(\omega)$ is conditioned by equation (1·5·4). This is an equation which holds identically for all modifications of $G(\omega)$, such as parameter variation, so that it will also hold for $\hat{G}(\omega)$:

$$\int_0^{2\pi} \log \hat{G}(\omega)\,d\omega = 0. \tag{2}$$

Combining (1) and (2) we obtain the fundamental estimation equation,

$$\frac{1}{2\pi}\int_0^{2\pi} \left[\frac{f(\omega)}{\hat{G}(\omega)} + \frac{1}{\alpha}\log \hat{G}(\omega)\right] d\omega = \text{min.} = \hat{v} \tag{3}$$

where α^{-1} is a Lagrangian multiplier.

If the minimisation (3) is carried out freely with respect to the *function* $G(\omega)$, then the calculus of variations gives

$$\left[\frac{\partial}{\partial G}\left[\frac{f}{G} + \frac{1}{\alpha}\log G\right]\right]_{G=\hat{G}} = 0 \tag{4}$$

or $\hat{G}(\omega)=\alpha f(\omega)$, a result which should scarcely surprise us. In general, however, minimisation may take place only with respect to a limited number of parameters.

For theoretical calculations it is convenient to work in terms of the periodogram, as we have done in equations (1)–(4), but for practical estimation a far preferable method is to use the correlogram [see equation (2·2·7)] so that the estimation equation is rather

$$\Sigma\,\hat{\gamma}_s C_s = \text{min.} = \hat{v} \tag{5}$$

where $[\hat{G}(\omega)]^{-1}=\Sigma\,\hat{\gamma}_s z^s$. The advantage is that we work in terms of a sum rather than an integral, and although the sum is theoretically an infinite one, the coefficients converge so quickly that it is usually sufficient to consider between 5 and 20 terms. Further, there is generally no difficulty in incorporating condition (2) at an early stage of the argument, so that the resultant minimisation with respect to parameters may be carried out freely.

Suppose, for instance, that $F(\omega)$ is rational in z, so that

$$F(\omega)=K\frac{\Pi\,(z-\beta)\,(z^{-1}-\beta)}{\Pi\,(z-\alpha)\,(z^{-1}-\alpha)}. \tag{6}$$

Since

$$\int_0^{2\pi}\log\,(z-\alpha)\,(z^{-1}-\alpha)\,d\omega=0 \qquad (|\,\alpha\,|<1) \tag{7}$$

we find

$$G(\omega)=\frac{\Pi\,(z-\beta)\,(z^{-1}-\beta)}{\Pi\,(z-\alpha)\,(z^{-1}-\alpha)} \tag{8}$$

so that γ_s is the coefficient of z^s in the Laurent expansion of

$$\frac{\Pi\,(z-\alpha)\,(z^{-1}-\alpha)}{\Pi\,(z-\beta)\,(z^{-1}-\beta)}. \tag{9}$$

Suppose that we consider, as example, a second order autoregression

$$x_t+a_1x_{t-1}+a_2x_{t-2}=\eta_t. \tag{10}$$

Then, by (5) and (9), we must minimise

$$(1+a_1^2+a_2^2)\,C_0+2\,(a_1+a_1a_2)\,C_1+2a_2C_2 \tag{11}$$

and so obtain readily the equations

$$\begin{aligned} C_1+\hat{a}_1C_0+\hat{a}_2C_1&=0\\ C_2+\hat{a}_1C_1+\hat{a}_2C_0&=0 \end{aligned} \tag{12}$$

which, on comparison with WOLD's equation (312), prove to be the classical estimation equations for the autoregressive scheme, as used by YULE. However, it is only for the autoregressive scheme that the least square estimates coincide with those that have been customary. Consider, for example, the first order moving average

$$x_t = \eta_t - b\,\eta_{t-1} \qquad (|b| < 1). \tag{13}$$

We find from (5), (9) that the expression to be minimised is

$$\frac{1}{1-b^2}\left[C_0 + 2\,b\,C_1 + 2\,b^2 C_2 + 2\,b^3 C_3 + \cdots\right]. \tag{14}$$

The estimate of b thus obtained is a function of the entire correlogram, although nearly all the weight falls on the earlier coefficients. It may easily have a variance only one quarter of that of the estimate obtained from C_0 and C_1 [see ref. 58].

2·6. *Variance of estimates*

In the case of independent variates least square estimates are usually consistent, and have certain optimum properties. WALD [ref. 52] has given a general set of sufficient conditions for the consistency and asymptotic optimality of maximum likelihood estimates which, when specialised to the present case (we assume normality for the moment), become that

(1) $$\frac{\partial^j}{\partial \theta^j}\left[\frac{1}{G(\omega)}\right]$$

exists and has a Fourier expansion ($j = 1, 2, 3$) in some θ interval around the true value and

(2) $$\int \left(\frac{G'}{G}\right)^2 d\omega \neq 0.$$

To these must be added the previous conditions that G and G^{-1} possess Fourier expansions. These conditions will in general cover also the case of non-normality although in this more general case the least-squares estimate is only asymptotically optimum among those estimates derived from linear relations in the autocovariances.

Let us now denote the residual sum of squares by U, $\dfrac{\partial U}{\partial \theta}$ by U_1, $\dfrac{\partial^2 U}{\partial \theta^2}$ by U_{11}, and the corresponding quantities for which θ assume its estimated value $\hat{\theta}$, by $\hat{U}$, $\hat{U}_1$, $\hat{U}_{11}$. WALD's conditions are sufficient to permit the expansion of $\hat{U}_1$ in a Taylor series

$$\hat{U}_1 = 0 = U_1 + (\hat{\theta} - \theta)\,U_{11} + 0(n^{-1}) \tag{1}$$

so that

$$\hat{\theta} - \theta \approx -\frac{U_1}{U_{11}}$$

and

$$\operatorname{var} \hat{\theta} = E(\hat{\theta} - \theta)^2 = E\left(\frac{U_1}{U_{11}}\right)^2 \approx \frac{E(U_1^2)}{[E(U_{11})]^2}. \tag{2}$$

Now, by equation (2·3·10) and identity (1·5·4) we have

$$E(U_1) = \frac{nv}{2\pi}\int_0^{2\pi} G \frac{\partial}{\partial\theta}\left(\frac{1}{G}\right) d\omega = -\frac{nv}{2\pi}\frac{\partial}{\partial\theta}\int \log G \, d\omega = 0 \tag{3}$$

$$E(U_{11}) = \frac{nv}{2\pi}\int_0^{2\pi} G \frac{\partial^2}{\partial\theta^2}\left(\frac{1}{G}\right) d\omega = \frac{nv}{2\pi}\int \left(\frac{G'}{G}\right)^2 d\omega \tag{4}$$

$$E(U_1^2) = \frac{2nv^2}{2\pi}\int_0^{2\pi} \left[G \frac{\partial}{\partial\theta}\left(\frac{1}{G}\right)\right]^2 d\omega = \frac{2nv^2}{2\pi}\int \left(\frac{G'}{G}\right)^2 d\omega. \tag{5}$$

Substituting (4) and (5) in (2) we can express the asymptotic variance of θ in terms of the spectral function as

$$\operatorname{var} \hat{\theta} = \frac{1}{\dfrac{n}{4\pi}\displaystyle\int \left(\frac{G'}{G}\right)^2 d\omega}. \tag{6}$$

For the multi-parameter case we can prove in the same way [see ref. 58] that the covariance matrix of $\hat{\theta}_1, \hat{\theta}_2 \ldots \hat{\theta}_p$ is asymptotically

$$\Delta = \left[\frac{n}{4\pi}\int_0^{2\pi}\left(\frac{\partial \log G}{\partial \theta_j}\right)\left(\frac{\partial \log G}{\partial \theta_k}\right) d\omega\right]^{-1}. \tag{7}$$

$$(j, k = 1, 2 \ldots p)$$

Further, $\hat{\theta}_1, \hat{\theta}_2 \ldots \hat{\theta}_p$ are uncorrelated with $\hat{v}$, which has variance $2v^2/n$.

An interesting special case is that in which the process is either an autoregression or a moving average, so that

$$G(\omega) = |1 + \theta_1 e^{-i\omega} + \theta_2 e^{-2i\omega} + \cdots + \theta_p e^{-pi\omega}|^{\pm 2} = |K(\omega)|^{\pm 2} \tag{8}$$

say. Substitution in (7) then gives

$$\Delta = \left[\frac{n}{2\pi}\int_0^{2\pi}\frac{e^{i(j-k)\omega}}{|K(\omega)|^2} d\omega\right]^{-1} \tag{9}$$

ndependently of whether the power in (8) is positive or negative, which ndicates an interesting symmetry between the two schemes.

Formula (7) holds generally, independently of the residual variate's istribution, since the estimates θ are functions of *ratios* of quadratic unctions, (easiest seen by noting that the estimates are unchanged if the stimation equation (2·5·5), is divided through by C_0).

It may be shown that the least square estimates are asymptotically ptimum in the sense that the total variance, $|\Delta|$, is a minimum, if we onfine ourselves to those estimates derived from second order functions f the observations. In the case of normally distributed residuals the ninimum is an absolute one.

2·7. *Tests of fit*

We shall now turn our attention to the test problem, the problem of leciding between a number of more or less distinct hypotheses. The NEYMAN-PEARSON test theory shows that, on a certain criterion, the statistic which liscriminates most efficiently between two specified hypotheses H_1, H_2, is the likelihood ratio. If the values of any of the parameters are unspecified then the appropriate statistic is the ratio of maximum likelihoods [refs. 35, 51, 54], at least asymptotically. This is equivalent to the ratio of minimised residual variances

$$\lambda = \frac{\hat{v}_2}{\hat{v}_1} \tag{1}$$

if the variates are normally distributed. In practice we have generally no idea of the variate distribution, but use λ anyway, on the grounds that

(a) least-square statistics are relatively easy to calculate

and (b) have a relatively simple distribution theory,

(c) few residual variates have a distribution which is radically non-normal,

and (d) even λ has certain limited optimum properties.

Despite statement (b) the calculation of the λ distribution is in general far from easy. We note one important class of exceptions, however, those cases for which hypothesis H_2 *includes* H_1. (Thus, for example, the hypothesis of a third order moving average includes those of a first or second order moving average, the hypothesis of stationarity includes that of periodicity.) Of course, when H_1 is merely a special case of H_2 we can no longer speak of discrimination, since the hypotheses are no longer mutually exclusive, as is tacitly assumed in most test theory. The statistic λ tests

whether the extra parameters of H_2 allow a significant improvement in fit, or, if we like, *tests the fit of H_1 relative to H_2*.

Suppose that hypothesis H_1 entails p undetermined parameters, while H_2 involves an extra q parameters, and so has $p+q$ altogether. We shall now briefly prove that

$$\psi^2 = (n-p-q)\log\frac{\hat{v}_1}{\hat{v}_2} \tag{2}$$

is, on hypothesis H_1, asymptotically distributed as χ^2 with q degrees of freedom [see refs. 55, 57, 53].

In equations (2·6·4), (2·6·5) we established the relation

$$E(U_1^2) = 2vE(U_{11}) \tag{3}$$

For the multiparameter case we can prove in exactly the same way

$$E(U_1U_1') = 2vE(U_{11}) \tag{4}$$

$$E(UU_{11}) = (n+2)vE(U_{11}) = \tfrac{1}{2}(n+2)E(U_1U_1') \tag{5}$$

where U_1 and U_{11} are now respectively the vector and matrix of first and second differential coefficients of U w.r.t. $\theta_1, \theta_2 \ldots \theta_p$. We shall denote the vectors of parameters and parameter estimates by θ, $\hat{\theta}$ respectively. Now, expanding $\log\hat{U}$ and $\dfrac{\partial}{\partial\hat{\theta}}\log\hat{U}$ in a Taylor series, we have

$$\log\hat{U} = \log U + \frac{U_1'}{U}(\hat{\theta}-\theta) + \tfrac{1}{2}(\hat{\theta}-\theta)'\left(\frac{U_{11}}{U} - \frac{U_1U_1'}{U^2}\right)(\hat{\theta}-\theta) + \cdots \tag{6}$$

$$0 = \frac{U_1}{U} + \left(\frac{U_{11}}{U} - \frac{U_1U_1'}{U^2}\right)(\hat{\theta}-\theta) + \cdots \tag{7}$$

where the terms will decrease regularly in powers of $n^{\frac{1}{2}}$, since $\hat{\theta}-\theta = 0(n^{-\frac{1}{2}})$. Now, by (7)

$$\hat{\theta}-\theta \approx -\left(U_{11} - \frac{U_1U_1'}{U}\right)^{-1}U_1 \tag{8}$$

and, substituting this approximate solution in (6), we have

$$\log U - \log\hat{U} \approx \tfrac{1}{2}U_1'(UU_{11} - U_1U_1')^{-1}U_1. \tag{9}$$

Equation (9) gives an approximate expression for the relative reduction in residual sum of squares due to fitting. Now, making use of the fact that $|E(U_{11})| \neq 0$, and of equations (4), (5), we find that

$$\log \frac{U}{\hat{U}} \approx \tfrac{1}{2}\, U_1' \,[E\,(U\,U_{11} - U_1\,U_1')]^{-1}\, U_1 = \frac{1}{n}\, U_1' \,[E\,(U_1\,U_1')]^{-1}\, U_1 \qquad (10)$$

so that $n \log \frac{U}{\hat{U}}$ would be asymptotically distributed as χ^2 with p d.f., if only U_1 were asymptotically normally distributed. This will in general be the case, however, as we see by an application of the BERNSTEIN extension of the Central Limit Theorem [ref. 7].

Suppose now that for the two hypotheses H_1, H_2 the min. residual sum of squares $\hat{U}$ has values $\hat{U}_p$, $\hat{U}_{p+q}$ respectively. Then $\log \frac{U}{\hat{U}_p}$, $\log \frac{U}{\hat{U}_{p+q}}$ are on hypothesis H_1 asymptotically distributed as χ^2 with p, $p+q$ d.f. respectively, p d.f. being common. Thus, by the partition theorem for χ^2 variates

$$n \log \frac{U}{\hat{U}_{p+q}} - n \log \frac{U}{\hat{U}_p} = n \log \frac{\hat{U}_p}{\hat{U}_{p+q}} \qquad (11)$$

is asymptotically distributed as χ^2 with q d.f. While not directly indicated in this derivation, the term n has been modified to $n-p-q$ in (2), to allow for "lost degrees of freedom" [cf. ref. 53].

The practical form of the test is, then, that a significantly high value of ψ^2 indicates bad fit, while a value in the neighbourhood of expectation indicates good fit, at least if we restrict ourselves to the alternatives permitted by H_2.

The simplest example of a ψ^2 statistic is that obtained on a comparison of two autoregressive schemes of different orders. For an autoregressive scheme of order p we find that

$$v = \frac{D_p}{D_{p-1}} \qquad (12)$$

where

$$D_p = \begin{vmatrix} C_0\, C_1 & \dots & C_p \\ C_1\, C_0 & \dots & C_{p-1} \\ \dots & \dots & \dots \\ C_p\, C_{p-1} & \dots & C_0 \end{vmatrix} \qquad (13)$$

so that for autoregressive schemes of order p, $p+q$

$$\lambda = \frac{\hat{v}_2}{\hat{v}_1} = \frac{D_{p+q}\, D_{p-1}}{D_{p+q-1}\, D_p} \qquad (14)$$

[see ref. 12] and

$$\psi^2 = -(n-p-q) \log \lambda \qquad (15)$$

may, for sufficiently large n, be tested as a χ^2 variate with q d.f. For other schemes than the autoregressive explicit expressions for the variance estimates do not exist [see eq. (2·5·14)], and these must be obtained by an iterative minimisation.

2·8. *Tests of fit with constant counterhypothesis*

The χ^2 test of the previous section measures the fit of H_1 *relative to the alternatives permitted by* H_2, *the counterhypothesis*. Now, in general the most desirable class of alternatives will be a very wide one — the class of all stationary processes, for instance, or the class of all stationary and purely nondeterministic processes. However, if H_2 is to be so comprehensive we see that we shall have some difficulty in calculating the corresponding variance estimate, v, since a hypothesis of the generality of those we have mentioned entails an *infinite* number of parameters. For example, if H_2 is the hypothesis that the process $\{x_t\}$ is stationary and purely nondeterministic, then

$$x_t = \eta_t + b_1\eta_{t-1} + b_2\eta_{t-2} + \cdots \tag{1}$$

and we may regard $(b_1, b_2 \ldots)$ as being distinct parameters of this general hypothesis. It is obvious that we cannot estimate all these parameters from a finite sample.

However, in this case we note that the sequence $b_1, b_2 \ldots$ must tend to zero, so that it is actually only the earlier coefficients which are of importance. Thus, it should in practice be possible to obtain a sufficient approximation to the actual process by estimating only a number of the earlier coefficients in (1), i.e., by graduating the series with a high order moving average.

Suppose that the graduation is of order k, and that the resultant least-squares estimate of the residual variance is $\hat{v}_0$. Suppose further that the fitting of a specified p-parameter hypothesis H leads to a variance estimate $\hat{v}$. Then, as before, for sufficiently large n *and* k

$$\psi^2 = (n-k)\log\left(\frac{\hat{v}}{\hat{v}_0}\right) \tag{2}$$

is asymptotically distributed as χ^2 with $k-p$ d.f. The advantage of (2) is that we now have a constant counterhypothesis, which is fairly sure to include all likely null hypotheses, and against which the fit of all such hypotheses may be tested.

This derivation of the test is, of course, an extremely superficial one. Caution is necessary in the application of (2), the most sensitive point being

the choice of the graduation order, k. For a fuller description and commentary see the original paper [ref. 57].

In practice, an autoregressive graduation is to be preferred to any other, as $\hat{v}_0$ may then be expressed directly in terms of sample functions [see eq. (2·8·12)] without the need to first explicitly solve for the fitted "parameters." Of course, if the process is to be representable as an autoregression, then $[F(\omega)]^{-1}$ must be expandable in a Fourier series, which is a restriction, but a minor one.

2·9. *Discriminatory tests*

For the general discriminatory test, the hypotheses H_1 and H_2 corresponding to the statistic

$$\lambda = \frac{\hat{v}_2}{\hat{v}_1} \tag{1}$$

bear no particular relation to one another. This increases the difficulty of coping with the unknown parameter values, and makes the whole treatment more approximate, in that we can only take account of differences in $\hat{v}_1$ and $\hat{v}_2$ of order $0(1)$, whereas in the previous two sections we considered differences of order $0(n^{-1})$.

Suppose that the minimised residual sum of squares for a hypothesis H is $\hat{U} = n\hat{v}$. Then, since $\hat{U}$ is a minimum w.r.t. $\theta_1, \theta_2 \ldots \theta_p$, $\hat{U}$ has a distribution which is asymptotically independent of the parameter estimates $\hat{\theta}_1, \hat{\theta}_2 \ldots \hat{\theta}_p$, whether H is the correct hypothesis or no [see 2·6]. Thus, $\hat{U}$ is asymptotically distributed as though the equality held

$$\hat{U} = \frac{n}{2\pi} \int_0^{2\pi} \frac{f(\omega)}{\overline{G}(\omega)}\, d\omega \tag{2}$$

where $\overline{G}(\omega)$ is the least-squares estimate of $G(\omega)$ *when* $f(\omega)$ *assumes its expectation value*. That is, $\hat{U}$ is asymptotically distributed as a certain linear function of the sample autocovariances. Thus, if $H(\omega)$ is the actual spectral function, then by (2·3·8) the r^{th} cumulant of $\hat{U}$ is

$$k_r \approx \frac{2^{r-1}(r-1)!\, n}{2\pi} \int_0^{2\pi} \left[\frac{H(\omega)}{\overline{G}(\omega)}\right]^r d\omega. \tag{3}$$

Similarly, the joint cumulants of $\hat{U}_1$ and $\hat{U}_2$ are given by

$$k_{rs} \approx \frac{2^{r+s-1}(r+s-1)!\, n}{2\pi} \int_0^{2\pi} \left[\frac{H(\omega)}{\overline{G}_1(\omega)}\right]^r \left[\frac{H(\omega)}{\overline{G}_2(\omega)}\right]^s d\omega. \tag{4}$$

Equation (4) gives us, at least in theory, a method of calculating the λ distribution, by using a bivariate type A series. This is not a very practical procedure, however, and we shall obtain a simpler one, at the price of another approximation.

Suppose that $G_2(\omega)$ corresponds to the null hypothesis, so that $H(\omega) \approx$ const. $\bar{G}_2(\omega)$, and suppose that n is so large that $\hat{U}_1$ and $\hat{U}_2$ may be considered as normally distributed. Then we find, with the help of a formula due to GEARY [ref. 20] that

$$\sqrt{\frac{n}{2}} \frac{1-\mu\lambda}{\sqrt{1-2\mu\lambda+\nu\lambda^2}} \tag{5}$$

is asymptotically normally distributed with zero mean and unit variance. The quantities μ and ν are given by

$$\mu = \frac{1}{2\pi}\int_0^{2\pi} \left[\frac{\bar{G}_2}{\bar{G}_1}\right] d\omega \qquad \nu = \frac{1}{2\pi}\int_0^{2\pi} \left[\frac{\bar{G}_2}{\bar{G}_1}\right]^2 d\omega \tag{6}$$

and so depend upon $H(\omega)$ and only upon $H(\omega)$, whose parameters must in practice be chosen in the neighbourhood of their estimated values.

2·10. *Harmonic components. The periodogram*

We have hitherto considered the estimation and testing only of a purely nondeterministic scheme, so the inclusion of a deterministic component is required to complete the treatment. In the main we shall restrict our attention to those deterministic components consisting of a finite number of sinusoidal terms,

$$\psi_t = \sum_{\nu=1}^{q} (A_\nu \sin \omega_\nu t + B_\nu \cos \omega_\nu t) \tag{1}$$

as being almost the only case of practical interest.

In this section we shall assume that the nondeterministic component is random, i.e., that $x_t = \psi_t + \eta_t$. Then we readily find that the least-squares estimates of A_ν, B_ν and ω_ν are approximately given by [see the next section for more detail]

$$\begin{gathered} \hat{A}_\nu = \frac{2}{n}\Sigma\, x_t \sin \hat{\omega}_\nu t \qquad \hat{B}_\nu = \frac{2}{n}\Sigma\, x_t \cos \hat{\omega}_\nu t \\ f(\hat{\omega}_\nu) = \frac{1}{n}[(\Sigma\, x_t \sin \hat{\omega}_\nu t)^2 + (\Sigma_t \cos \hat{\omega}_\nu t)^2] = \max. \end{gathered} \tag{2}$$

so that the least square estimates of the ω_ν's are located at the peaks of the periodogram $f(\omega)$, and the magnitudes of these peaks provide the least-square estimates of the corresponding amplitudes, $A_\nu^2 + B_\nu^2$. It may be shown that these estimates have least variance of any yielded by second order functions of the observations, despite the exceptional character of the ω_ν estimate [see 2·12].

We see thus that the SCHUSTER periodogram analysis, of late years regarded with so much uncertainty, is actually the appropriate technique for the location and estimation of periodic components. That it should ever have fallen into disfavour is a result of mistaken (although extremely understandable!) attempts to use it for purposes for which it was not intended [see 2·12]. It is notable that the periodogram, which with a 55 year history behind it may fairly be claimed to be the time series analyst's oldest weapon, has survived unchanged and shown itself scarcely capable of improvement. The same is true of the classical periodogram test, which we shall now describe, derived by FISHER (1929).

If the number of observations is odd and

$$y_j = f\left(\frac{2\pi j}{n}\right) \qquad \left(j = 1, 2 \ldots m;\; m = \frac{n-1}{2}\right) \tag{3}$$

then FISHER takes the greatest of the ordinates y_j, (let it be y_a) and shows that the probability of obtaining a value of

$$g = y_a / \sum_1^m y_j \tag{4}$$

greater than that observed is

$$\sum_{j=1} (-)^{j-1}\, j \binom{m}{j} (1 - j\,g)^{m-2} \tag{5}$$

where the summation is continued over all terms in which $(1 - j\,g)$ is positive. The assumption is that the observations are distributed normally and independently about zero.

We see how radically the test differs from the usual type: if it were so that one "degree of freedom" were lost for each of the three estimated parameters A, B and ω, then g would have an expected value $3/n$, whereas a calculation shows the expectation to be of order $\frac{2}{n} \log \frac{n}{2}$ (consider the extreme values of the rectangularly distributed variate $\exp[-y/E(y)]$). However, as soon as a genuine harmonic component appears the minimisation w.r.t. ω "stabilises", and we shall see later that the relative reduction in residual sum of squares due to fitting is in this case really of the order $3/n$.

At it stands, FISHER's test is limited in three respects:

(a) Only the largest peak may be tested.

(b) The only null hypothesis is that of independence.

(c) The only ordinates considered are the equidistant ones (3).

The first two restrictions are easily remedied. First, suppose that we wish to test the second largest ordinate, y_b. This will be the case only when y_a has been found significant, so that we must ipso facto modify our hypothesis to include a harmonic term with an ω value in the neighbourhood of y_a's. However, the distribution of the remaining ordinates will be little affected by this, and

$$g' = \frac{y_b}{\sum y_j - y_a} \tag{6}$$

will have approximately the distribution (5) with m replaced by $m-1$. Similarly for the third greatest, etc.

Suppose now that we wish to adopt the null hypothesis that the scheme has a known continuous spectrum, $F(\omega)$. This has the approximate effect of altering $f(\omega)$'s scale factor by $F(\omega)$, [see (2·2·6)], so that if we define

$$y_j = f\left(\frac{2\pi j}{n}\right) \Big/ F\left(\frac{2\pi j}{n}\right) \tag{7}$$

then g will have the same definition and almost the same distribution as before.

The third limitation is not so easily removed, but is on the other hand less serious. If a harmonic term falls midway between two ordinates of the grid (3) then its amplitude will be reduced by roughly $4/\pi^2 \approx 0.41$ at these ordinates, which may or may not be enough to make it appear nonsignificant. Some headway may be made by considering integrated values of the periodogram over an ω interval, see 2·12.

2·11. *The composite case*

Let us now consider a sample from the general stationary process (with the single restriction that the purely nondeterministic component of the spectrum be nonzero in $(0, 2\pi)$). Since the deterministic and nondeterministic components are additive, the residual sum of squares must be expressible as in 2·2, except that x_t will be replaced by $x_t - \psi_t$. Introducing this modification in eq. (2·2·8) we find that

$$U = \frac{1}{2\pi} \int_0^{2\pi} \frac{n f(\omega) - 2A(\omega)X(-\omega) + A(\omega)A(-\omega)}{G(\omega)} d\omega \tag{1}$$

where

$$X(\omega)=\sum_1^n x_t e^{i\omega t} \qquad A(\omega)=\sum_1^n \psi_t e^{i\omega t}. \tag{2}$$

Suppose that $G(\omega)$ and $A(\omega)$ involve parameters $\theta_1, \theta_2 \dots \theta_p$, whose least-square estimates $\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_p$ are obtained by minimising U of (1). Then, using the same methods as in 2·6, we find the covariance matrix of $\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_p$ to be asymptotically

$$\Delta = \left[\frac{n}{4\pi}\int_0^{2\pi}\left(\frac{\partial \log G(\omega)}{\partial \theta_j}\right)\left(\frac{\partial \log G(\omega)}{\partial \theta_k}\right) d\omega + \frac{1}{2\pi v}\int_0^{2\pi}\frac{A^{(j)}(\omega)\, A^{(k)}(\omega)}{G(\omega)}\, d\omega\right]^{-1}$$

$$(j,\ k=1,\ 2,\ \dots\ p) \tag{3}$$

where

$$A^{(j)}(\omega)=\frac{\partial}{\partial \theta_j} A(\omega). \tag{4}$$

Of course, formula (3) and its derivation break down if Δ^{-1} should be singular, so that Δ does not exist. Such would be the case, for example, if we should fit a harmonic component where none existed (for then the $\dfrac{\partial}{\partial \omega_\nu}$ column in Δ^{-1}, which is proportional to the amplitude of the component, would be zero).

The tests of fit of 2·7–2·8 may be extended to the present case, again under the restriction that Δ exist for all hypotheses considered.

The case of greatest interest is again that for which

$$\psi(t)=\sum_{\nu=1}^{q}(A_\nu \sin \omega_\nu t + B_\nu \cos \omega_\nu t). \tag{5}$$

Substituting (5) in (1), and using the relations

$$\Sigma \sin^2 \omega t \approx \Sigma \cos^2 \omega t \approx \frac{n}{2}$$

$$\Sigma \sin \omega t \cos \omega t \approx 0$$

$$\Sigma \begin{pmatrix}\sin \omega_\mu t\\ \cos \omega_\mu t\end{pmatrix}\begin{pmatrix}\sin \omega_\nu t\\ \cos \omega_\nu t\end{pmatrix} \approx 0 \qquad (\mu \neq \nu) \tag{6}$$

we find, after some reduction, that

$$U \approx \frac{n}{2\pi}\int_0^{2\pi}\frac{f(\omega)}{G(\omega)}\, d\omega - 2\sum_\nu \frac{(A_\nu a_\nu + B_\nu b_\nu)}{G(\omega_\nu)} + \frac{n}{2}\sum_\nu \frac{(A_\nu^2 + B_\nu^2)}{G(\omega_\nu)} \tag{7}$$

where

$$a_\nu = \Sigma x_t \sin \omega t \qquad b_\nu = \Sigma x_t \cos \omega t. \tag{8}$$

The estimation equations for $\hat{A}_\nu$, $\hat{B}_\nu$ and $\hat{\omega}_\nu$ are found from (7) to be

$$\hat{A}_\nu = \frac{2}{n} a_\nu \qquad \hat{B}_\nu = \frac{2}{n} b_\nu$$

$$\frac{\partial}{\partial \omega_\nu} \left[\frac{a_\nu^2 + b_\nu^2}{G(\omega_\nu)} \right] = 0 \tag{9}$$

so that the harmonic components are chosen from the greatest peaks in $f(\omega)/G(\omega)$, although the amplitudes are estimated directly from $f(\omega)$.

The complete system of estimation equations is

$$\hat{A}_\nu = \frac{2\,a_\nu}{n} \qquad \hat{B} = \frac{2\,b_\nu}{n} \qquad \frac{f(\hat{\omega}_\nu)}{\hat{G}(\hat{\omega}_\nu)} = \max. \qquad (\nu = 1, 2, \ldots g)$$

$$\frac{n}{2\pi} \int_0^{2\pi} \frac{f(\omega)}{\hat{G}(\omega)} d\omega - 2 \sum_\nu \frac{f(\hat{\omega}_\nu)}{\hat{G}(\hat{\omega}_\nu)} = \min. \tag{10}$$

$$\int_0^{2\pi} \log \hat{G}(\omega)\, d\omega = 0,$$

the maximum and minimum being taken with respect to ω_ν and $(\theta_1, \theta_2 \ldots \theta_p)$ respectively. Applying (3), or working direct from (7), we find that $\hat{A}_\mu$, $\hat{B}_\mu$, $\hat{\omega}_\mu$ are uncorrelated with $\hat{A}_\nu$, $\hat{B}_\nu$, $\hat{\omega}_\nu$ $(\mu \neq \nu)$, or with the parameter estimates of $G(\omega)$, $\hat{\theta}_1, \hat{\theta}_2 \ldots \hat{\theta}_p$. These latter have the same covariance matrix as before [eq. (2·6·7)], while the covariance matrix of $\hat{A}_\nu$, $\hat{B}_\nu$, $\hat{\omega}_\nu$ is

$$\frac{2\,v\,G(\omega_\nu)}{n} \begin{bmatrix} 1 & 0 & \dfrac{-n B_\nu}{2} \\ 0 & 1 & \dfrac{n A_\nu}{2} \\ \dfrac{-n B_\nu}{2} & \dfrac{n A_\nu}{2} & \dfrac{n^2}{6}(A_\nu^2 + B_\nu^2) \end{bmatrix}^{-1}. \tag{11}$$

The least-squares estimate of a constant mean is given approximately by $\bar{x} = \frac{1}{n} \Sigma x_t$. This is uncorrelated with any of the other parameters and has variance $v G(o)/n$. Regarding the practical solution of the equation system (10), see ref. 56.

2·12. *Miscellaneous periodogram results*

One characteristic of the periodogram is almost too well-known to need mention, namely, that while it gives an estimate of the spectral intensity $F(\omega)$ at a continuous part of the spectrum, this estimate is inconsistent. This may be quickest seen in the following manner:

$$f(\omega) = \frac{1}{n}[(\Sigma x_t \sin \omega t)^2 + (\Sigma x_t \cos \omega t)^2] = \frac{1}{n}\sum_1^n \sum_1^n x_s x_t \cos \omega (s-t)$$

$$\approx \sum_{-n}^{n} C_s e^{i\omega s} \qquad (1)$$

so that

$$E[f(\omega)] = \sum_{-n}^{n} \left[1 - \frac{|s|}{n}\right] \varrho_s e^{i\omega s} \qquad (2)$$

$$\text{var}\,[f(\omega)] \approx [G(\omega)]^2 \,\text{var}\,[f^{(\eta)}(\omega)] = 2[G(\omega)]^2 v^2. \qquad (3)$$

Thus, while $E[f(\omega)]$ is equal to $F(\omega)$, apart from a small bias vanishing with increasing n, var $[f(\omega)] = 0\,(1)$, so that the coefficient of variation does not vanish as $n \to \infty$. This means in practice that the periodogram presents a wildly irregular appearance, suggesting little or nothing to the eye.

That a harmonic term (corresponding to an *infinite* $F(\omega)$ ordinate) can be estimated consistently depends upon the fact that the amplitude of such a harmonic is estimated by $(4/n)\, f(\omega)$, and has coefficient of variation $0\,(n^{-\frac{1}{2}})$.

The root of the trouble is the fact that the periodogram is an empirical frequency function — it describes the distribution of the "energy" of the observed fluctuations among the different frequencies of oscillation (from which follows, incidentally, that it is incorrect to set up a periodogram graph with the period of oscillation as abscissa, as is sometimes done, without scaling the ordinates by the Jacobian of the transformation from frequency to period). The phenomenon of inconsistency is one characteristic of all empirical frequency functions of a continuous variate, namely, that while the total frequency in any interval converges to a more or less determinate quantity with increasing sample size, the point density of observations never does so, unless the point should be one of infinite probability density.

We have seen that in the least square approach to the inference problem, it has been necessary to consider the periodogram as such only when estimating harmonic terms, otherwise we have worked with periodogram averages (in actual fact, what is the same thing, linear functions of the earlier autocovariances). However, in many applications it would be useful

to be able to form a direct estimate of the spectrum, and various methods of smoothing the periodogram have been suggested in order to bring this about.

BARTLETT [ref. 3] has proposed the estimate provided by the Fourier transform of the truncated correlogram

$$f_B(\omega) = \sum_{-r}^{r} C_s e^{i\omega s} \tag{4}$$

[cf. (1)] where r has usually a value between 15 and 30, depending upon circumstances. This estimate is justified by the fact that the latter part of the correlogram of a purely nondeterministic series contains more random variation than useful information (which is why the earlier coefficients are always heaviest weighted in the least-square estimation equations). Alternatively, (4) may be regarded as being derived as the average of the n/r periodograms calculated for n/r series of length r. By eq. (2·3·10)

$$k_j[f_B(\omega)] = \left(\frac{2}{n}\right)^{j-1} \frac{(j-1)!}{2\pi} \int_0^{2\pi} \left[\frac{z^{-r} - z^{r+1}}{1-z} F(\omega)\right]^j d\omega \tag{5}$$

so that the variance is $0\,(r/n)$.

DANIELL [ref. 11] has suggested smoothing the periodogram by a moving average, so that if we average over $\omega \pm h$

$$f_D(\omega) = \sum_{-n}^{n} C_s e^{i\omega s} \frac{\sin(sh)}{sh}. \tag{6}$$

As h increases the variance of the estimate decreases, but so also does the frequency resolvability. This provides an illustration of the uncertainty principle enunciated by GRENANDER [ref. 22] stating that the product of errors in the estimates of amplitude and frequency has a certain lower bound.

It is perhaps unnecessary to add that both BARTLETT's and DANIELL's modifications are designed to improve the direct estimation of a *continuous* spectrum; they will both lead to serious underestimation of the amplitude of any harmonic component.

The integrated periodogram is an empirical distribution function, and so will give consistent estimates of the spectrum. Thus, by eq. (2·3·10) the cumulants of

$$f(\omega_1, \omega_2) = \frac{n}{2\pi} \int_{\omega_1}^{\omega_2} f(\omega)\, d\omega \quad (0 \leqq \omega_1 \leqq \omega_2 \leqq \pi) \tag{7}$$

are asymptotically given by

$$k_j = \frac{2^{j-1}(j-1)!\,n}{\pi} \int\limits_{\omega_1}^{\omega_2} [F(\omega)]^j \, d\omega . \tag{8}$$

From this formula we can calculate the cumulants of $f_D(\omega)$, if only h is not too small.

We shall conclude this section by calling attention to a peculiarity of $\hat{\omega}$, the least-squares estimate of the frequency of a harmonic component. The majority of parameter estimates have variance $0(n^{-1})$, but $\hat{\omega}$ has variance $0(n^{-3})$ or $0(1)$, depending upon whether the component actually exists or no. For if there is in reality no harmonic term, then it is clear that the periodogram can have a peak anywhere, while if the component exists, equation (11) of the previous section shows that $\hat{\omega}$ has variance $0(n^{-3})$.

2·13. *Multiple series*

Practical problems far more often require the analysis of several simultaneous and related series than of a single one. All of the preceding least-squares theory may be readily generalised to this case, although the multiplicity does introduce new features (e.g., "identifiability"). In this section we shall very briefly consider the least square treatment of a purely nondeterministic multiple series.

Suppose that we have r variates, $x_t = (x_{1t}, x_{2t} \ldots x_{rt})$ which are linearly dependent upon r mutually independent series of residual variates,

$$\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t} \ldots \varepsilon_{rt}),$$

so that

$$x_t = B(T)\varepsilon_t \tag{1}$$

Here $B(T)$ is an $r \times r$ matrix whose elements are functions of the operator T, defined by

$$Tu_t = u_{t-1} \tag{2}$$

Only positive powers of T are assumed to occur, i.e., x_t is a function of past values of ε_t alone. If $|B(z)|$ has no zeroes on or within the unit circle, then there is also an autoregressive representation of the multiple process:

$$\varepsilon_t = [B(T)]^{-1}x_t = A(T)x_t \tag{3}$$

say.

We shall now define the theoretical and empirical autocovariances and spectral functions:

$$\varrho_{jk}(s) = E[x_{j,\,t+s}\, x_{k\,t}] \qquad F(\omega) = (F_{jk}(\omega)) \tag{4}$$

$$F_{jk}(\omega) = \sum_{-\infty}^{+\infty} \varrho_{jk}(s)\, e^{i\omega s} \qquad C_{jk}(s) = \frac{1}{n}\sum_{t=1}^{n-s} x_{j,\,s+t}\, x_{kt} \tag{5}$$

$$f(\omega) = (f_{jk}(\omega)). \qquad f_{jk}(\omega) = \sum_{-n}^{n} C_{jk}(s)\, e^{i\omega s} \tag{6}$$

We readily find by direct methods from (1) and (3) that

$$F(\omega) = B(e^{i\omega})\, B'(e^{-i\omega}) = [A'(e^{-i\omega})\, A(e^{i\omega})]^{-1} \tag{7}$$

if we assume that the B coefficients have been chosen so that the residuals have unit variance. The least square estimate of the matrix of spectral functions $F(\omega)$ is yielded by the relation

$$\frac{1}{2\pi}\int_0^{2\pi} \log|\hat{F}|\, d\omega + \frac{1}{2\pi}\int_0^{2\pi} \mathrm{tr}\,[f\hat{F}^{-1}]\, d\omega = \min. \tag{8}$$

and the resulting parameter estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_p$ have asymptotic covariance matrix

$$\left[\frac{n}{4\pi}\int_0^{2\pi} \mathrm{tr}\left[\frac{\partial F}{\partial \theta_j} F^{-1} \frac{\partial F}{\partial \theta_k} F^{-1}\right] d\omega\right]^{-1} \quad (j, k = 1, 2 \ldots p) \tag{9}$$

if the variates are normally distributed. A feature of the multivariate case is that (9) may not always be extended to the case of non-normal variates.

As before, if hypotheses H_1 and H_2 involve respectively p and $p+q$ parameters, and H_2 includes H_1, then

$$\psi^2 = \left(n - \frac{p+q}{r}\right) \log \frac{\hat{v}_1}{\hat{v}_2} \tag{10}$$

is asymptotically distributed as χ^2 with q degrees of freedom. Here

$$\hat{v}_j = e^{\frac{1}{2\pi}\int_0^{2\pi} \log|\hat{F}_j|\, d\omega} \qquad (j = 1, 2) \tag{11}$$

where $\hat{F}_1$, $\hat{F}_2$ are two least square estimates of $F(\omega)$.

The following supplementary references may be given for topics not covered here: for the more recent theory of stationary processes, refs. 13–17 for treatments of inference problems in the continuous case, refs. 2, 21, for discrete series with discontinuous variates, ref. 18, and for a test of fit for such series, ref. 4.

References.

J. BARTELS (1935): *Zur Morphologie geophysikalischer Zeitfunktionen.* Sitz.-ber. Preuss.-Akad. Wiss., Phys.-Math., Berlin, No. 30.

W. H. BEVERIDGE (1921): *Weather and harvest cycles.* Econ. J., *31*, 429.

H. BOHR (1925): *Zur Theorie der fast periodischen Funktionen.* Acta Mathem., *45*, 29.

B. BRUNO (1938): Die Wasserstandsschwankungen des Wenersees. *Geografiska Annaler*, *20*, 308.

H. BRUNS (1911): *Über die Analyse periodischer Vorgänge.* Astron. Nachr., *188*, 55.

D. BRUNT (1913): *The combination of observations.* 2nd ed. Cambridge.

—— (1937): *Climatic cycles.* Geogr. J., *89*, 214.

H. BURKHARDT (1904): *Trigonometrische Interpolation.* Enzykl. Math. Wiss., Bd II, Teil *1*, 1, 642.

C. H. D. BUYS-BALLOT (1847:) *Les changements périodique de température.* Utrecht.

F. P. CANTELLI (1916): *La tendenza ad un limite nel senso del calcolo delle probabilità.* R. circ. matem. Palermo, *41*, 191.

H. S. CARSLAW (1930): *Introduction to the theory of* FOURIER's *series and integrals.* 3rd ed. London.

J. I. CRAIG (1916): *A new method of discovering periodicities.* Month. Not. R. Astron. Soc., *76*, 493.

H. CRAMÉR (1927): *Sannolikhetskalkylen och några av dess användningar.* Stockholm.

—— (1928): *On the composition of elementary errors*, I, II. Skandinavisk Aktuarietidskift, *11*, 13 and 141.

—— (1937): *Random variables and probability distributions.* Cambridge.

W. DODGSON, see B. STEWART.

G. ELFVING (1937): *Zur Theorie der* MARKOFF*schen Ketten.* Acta Soc. Sci. Fennicæ (A), *2*, no. 8.

R. A. FISHER (1929): *Tests of significance in harmonic analysis.* Proc. Roy. Soc. (A), *125*, 54.

M. FRÉCHET (1937): *Généralités sur les probabilités. Variables aléatoires.* Paris.

R. FRISCH (1933): *Propagation problems and impulse problems in dynamic economics.* Economic essays in honour of GUSTAV CASSEL. London.

F. HAUSDORFF (1923): *Momentprobleme für ein endliches Intervall.* Mathem. Zeitschr., *16*, 220.

G. HERGLOTZ (1911): *Über Potenzreihen mit positivem, reellen Teil im Einheitskreis.* Ber. Leipzig Math.-Phys., *63*, 501.

M. G. KENDALL, see G. U. YULE.

A. KHINTCHINE (1932): *Sulle successioni stazionarie di eventi.* Giorn. Ist. Ital. Attuari, *3*, 267.

A. Khintchine (1933): *Über stationäre Reihen zufälliger Variablen.* Rec. mathém. Moscou, *40*, 124.

—— (1934): *Korrelationstheorie der stationären stochastischen Prozesse.* Mathem. Ann., *109*, 604.

A. Kolmogoroff (1931): *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung.* Math. Ann., *104*, 415.

—— (1933): *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Ergebn. Math. u. Grenzgeb., *2*, no. 3.

G. Kowalewski (1909): *Einführung in die Determinantentheorie.* Leipzig.

T. Lindblad (1937): *Zur Theorie der Korrelation bei mehrdimensionalen zufälligen Variablen.* Acta Soc. Sci. Fennicæ (A), *2*, no. 10.

E. Lundberg (1937): *Studies in the theory of economic expansion.* Stockholm.

P. M. Marples (1932): *Linear difference equations. An elementary treatment.* J. Inst. Actuaries, *63*, 404.

R. v. Mises (1931): *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik.* Leipzig und Wien.

G. Myrdal (1933): *The cost of living in Sweden 1830—1930.* Stockholm.

C. W. B. Normand (1932): *Present position of seasonal weather forecasting.* Q. J. Roy. Meteor. Soc., *58*, 3.

S. Oppenheim (1909): *Über die Bestimmung der Periode einer periodischen Erscheinung nebst Anwendung auf die Theorie der Erdmagnetismus.* Wien. Sitz.-Ber. (2a), *118*, 823.

G. Robinson, see E. T. Whittaker.

V. Romanovsky (1932): *Sur la loi sinusoïdale limite.* R. circ. matem. Palermo, *56*, 22.

—— (1933): *Sur une généralisation de la loi sinusoïdale limite.* R. circ. matem. Palermo, *57*, 130.

C. F. Roos (1934): *Dynamic economics.* Bloomington, Ind.

A. Schuster (1898): *On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena.* Terr. Magn., *3*, 13.

—— (1900): *The periodogram of the magnetic declination as obtained from the records of the Greenwich Observatory during the years 1871—1895.* Trans. Cambridge Phil. Soc., *18*, 107.

E. Slutsky (1925): *Über stochastische Asymptoten und Grenzwerte.* Metron, *5*, 3.

—— (1927): *The summation of random causes as the source of cyclic processes* (Russian with an English summary). Problems of economic conditions, *3*, no. 1.

—— (1934): *Alcune applicazioni dei coefficienti di* Fourier *all' analisi delle funzioni aleatorie stazionarie.* Giorn. Ist. Ital. Attuari, *5*, 435.

—— (1937): *The summation of random causes as the source of cyclic processes.* Econometrica, *5*, 105.

N. Smirnoff (1936): *Sur la distribution de* ω^2. C. R. Acad. Sci. Paris, *202*, 449.

B. Stewart and W. Dodgson (1879): *Preliminary report to the Committee on Solar Physics on a method of detecting the unknown inequalities of a series of observations.* Proc. Roy. Soc., *29*, 106.

B. STEWART and W. DODGSON (1879): *Preliminary report to the Committee on Solar Physics on the evidence in favour of the existence of certain short periods common to solar and terrestrial phenomena.* Proc. Roy. Soc., *29*, 303.

K. STUMPFF (1927): *Analyse periodischer Vorgänge.* Berlin.

—— (1936): *Über die Zufallswahrscheinlichkeit von Periodizitäten in Beobachtungsreihen.* Veröff. Meteor. Inst. Univ. Berlin, *1*, no. 2.

—— (1937): *Grundlagen und Methoden der Periodenforschung.* Berlin.

J. TINBERGEN (1937): *An econometric approach to business cycle problems.* Paris.

L. H. C. TIPPET (1927): *Random sampling numbers.* London.

SIR G. WALKER (1931): *On periodicity in series of related terms.* Proc. Roy. Soc. (A), *131*, 518.

E. T. WHITTAKER (1911): *On the law which governs the variations of SS Cygni.* Month. Not. R. Astron. Soc., *71*, 686.

E. T. WHITTAKER and G. ROBINSON (1926): *The calculus of observations.* 2nd ed. London.

N. WIENER (1930): *Generalized harmonic analysis.* Acta Mathem., *55*, 117.

H. WOLD (1935): *A study on the mean difference, concentration curves and concentration ratio.* Metron, *12*, no. 2, 39.

—— (1936): *On quantitative statistical analysis.* Skandinavisk Aktuarietidskrift, *19*, 281.

G. U. YULE (1921): *On the time-correlation problem, with especial reference to the variate-difference correlation method.* J. Roy. Stat. Soc., *84*, 497.

—— (1926): *Why do we sometimes get nonsense-correlations between time-series? — A study in sampling and the nature of time-series.* J. Roy. Stat. Soc., *89*, 1.

—— (1927): *On a method of investigating periodicities in disturbed series, with special reference to* WOLFER's *sunspot numbers.* Trans. Roy. Soc. (A), *226*, 267.

G. U. YULE and M. G. KENDALL (1937): *An introduction to the theory of statistics.* London.

Appendix references.

1. ANDERSON, R. L.: Ann. Math. Stat., *13* (1942) 1.
2. BARTLETT, M. S.: J. Roy. Stat. Soc., Ser. B, *8* (1946) 27.
3. —— : Nature, *161* (1948) 686.
4. —— : Proc. Cambr. Phil. Soc., *47* (1950) 86.
5. BAUMOL, in association with R. TURVEY: Economic dynamics. An introduction. New York 1951.
6. BENTZEL, R. and WOLD, H.: Skand. Aktuarietidskrift, *29* (1946) 95.
7. BERNSTEIN, S.: Math. Ann., *97* (1927) 1.
8. CRAMÉR, H.: Ann. Math., *41* (1940) 215.
9. —— : Ark. Mat. Astr. Fys., *28 B* (1942) no. 12.
10. —— : Mathematical methods of statistics. Uppsala 1945, Princeton 1946.

11. Daniell, P. J.: J. Roy. Stat. Soc., Ser. B, 8 (1946) 89.
12. Dixon, W. J.: Ann. Math. Stat., *14* (1944) 119.
13. Doob, J. L.: Trans. Amer. Math. Soc., *42* (1937) 107.
14. —— : Trans. Amer. Math. Soc., *44* (1938) 87.
15. —— : Ann. Math. Stat., *15* (1944) 229.
16. —— : Proc. Berkeley Symp. Math. Statistics and Prob. 1946, 303—343. Berkeley 1949.
17. —— : Stochastic processes. New York 1953.
18. Feller, W.: Probability theory and its applications, I. New York 1950.
19. Frisch. R.: Nordisk Stat. Tidskr., *8* (1928) 36.
20. Geary, R. C.: J. Roy. Stat. Soc., *93* (1930) 442.
21. Grenander, U.: Ark. Matem., *1* (1950) 195.
22. —— : Ark. Matem., *1* (1951) 503.
23. —— : Ark. Matem., *1* (1951) 555.
24. Gurland, J.: Cowles Comm. Discussion Papers. Statistics, no. 354 (1950).
25. Hopf, E.: Ergebn. Math. *5*, no. 2 (1937).
26. Karhunen, K.: Ann. Acad. Sci. Fennicae, Ser. A, 1. Math. Phys., no. 37 (1947).
27. —— : C. R. Dixième Congr. Math. Scandinaves 1946, 320—324. Copenhagen 1947.
28. Kendall, M. G.: J. Roy. Stat. Soc. *108* (1945) 93.
29. —— : The advanced theory of statistics, II. London 1946.
30. Kolmogoroff, A. N.: Giorn. Ist. Ital. Attuari, *4* (1933) 83.
31. —— : C. R. Acad. Sci. Paris, *208* (1939) 2043.
32. —— : Bull Acad. Sci. (Nauk) URSS, Ser. Math., *5* (1941) 3.
33. —— : Bull. State Univ. Moscow, Ser. Math., *2*, no. 6 (1941).
34. Koopmans, T.: Ann. Math. Stat., *13* (1942) 14.
35. Lehmann, E. L. and Stein, C.: Ann. Math. Stat., *19* (1948) 495.
36. Leipnik, R. B.: Ann. Math. Stat., *18* (1947) 80.
37. Lukomski, J.: Ann. Math. Stat., *10* (1939) 236.
38. Madow, W. G.: Ann. Math. Stat., *16* (1945) 308.
39. Mann, H. B.: and Wald, A.: Econometrica, *11* (1943) 173.
40. Moyal, J.: J. Roy. Stat. Soc., Ser. B, *11* (1949) 150.
41. Quenouille, M. H.: Ann. Math. Stat., *20* (1949) 561.
42. Robbins, H.: Ann. Math. Stat., *19* (1948) 360.
43. Rubin, H.: Ann. Math. Stat., *16* (1945) 211.
44. Slutsky, E.: Actualités Sci. Ind., *738* (1938) 35.
45. Szegö, G.: Math. Zeitschr., *6* (1920) 167.
46. —— : Math. Zeitschr., *9* (1921) 167.
47. Tinbergen, J.: Business cycles in the United States of America. Geneva 1939.
48. —— : Rev. Econ. Stud., *7* (1940) 73.
49. Tintner, G.: Econometrics. New York 1952.
50. —— : Mathematics and statistics for economists. New York [forthcoming].
51. Wald, A.: Trans. Amer. Math. Soc., *54* (1943) 426.

WALD, A.: Ann. Math. Stat., *19* (1948) 40.

WALKER, A.: J. Roy. Stat. Soc., Ser. B, *14* (1952) 117.

WHITTLE, P.: Hypothesis testing in time-series analysis. Uppsala 1951.

——: Skand. Aktuarietidskr., *35* (1952) 48.

——: Trab. Estadist., *3* (1952) 43.

——: Biometrika, *39* (1952) 309.

——: Ark. Matem., *2* (1953) 423.

——: J. Roy. Stat. Soc., Ser. B, *15* (1953) 125.

WIENER, N.: Extrapolation, interpolation, and smoothing of stationary time series. New York 1949.

WOLD, H.: Skand. Aktuarietidskr., *21* (1938) 208.

——: Ann. Math. Stat., *19* (1948) 558.

——: Econometrica, *17*, Suppl. (1949) 1.

——, in association with JURÉEN, L.: Demand analysis. A study in econometrics. Stockholm 1952, New York 1953.

ZASUHIN, V.: C. R. (Doklady) Acad. Sci. URSS, New ser., *33* (1941) 435.

INDEX.